

**University College London**

Division of Medicine

School of Life and Medical Sciences

Multivalent ligand recognition by

Pentraxins

A submission for the degree of Doctor of  
Philosophy

**Peter Hughes**

July 2015

## **Declaration**

I, Peter Hughes, confirm that the work presented in this thesis is my own. Where I have included material from other sources, I confirm that this has been indicated within the thesis.

Signed

---

Peter Hughes

## Abstract

The pentraxins, serum amyloid P component (SAP) and C-reactive protein (CRP) are target proteins for the development of treatments for amyloidosis and ischaemic injury, respectively, in humans. This study reports the first multivalent ligands capable of targeting all five SAP binding sites simultaneously. Ligands presenting five or ten D-proline headgroups and composed of five peptide-glycol linkers emanating from  $\epsilon$ -N-substituted lysine residues on a central cyclic peptide core were synthesised by solid phase peptide synthesis. The sub-nanomolar,  $\sim 250\text{pM}$ , binding affinity approximated by Isothermal Titration Calorimetry (ITC) for one decavalent ligand is the strongest affinity for an SAP binding ligand currently known and stronger than the affinity of SAP binding to amyloid deposits. X-ray crystallography and mass spectrometry shows the decavalent ligands non-covalently cross-linking two SAP pentamers, in the same manner observed for lead drug candidate CPHPC, but with increased affinity.

In addition, the binding of SAP with N-acetyl D-proline has been investigated by x-ray crystallography. Using a  $1.5\text{\AA}$  resolution structure the exact interaction of the headgroup used in CPHPC, penta- and decavalent ligands, was investigated. The results show potential for an electrostatic interaction between the carbonyl oxygen of acetyl from the ligand and the side chain amide of Gln148, which has not previously been considered.

Applications of multivalent binding are still emerging; in this study, bivalent ligand BPC8 was used as an additive to crystallise CRP from the Rat (rCRP) in non-covalently cross-linked decameric complexes. Previous x-ray crystallography studies have failed due to extreme radiation sensitivity of the crystals produced. This problem has been overcome with a complete dataset obtained from a single crystal at  $3.2\text{\AA}$ . No inter-protein contacts are seen between pentamers in the decamer complex, therefore the use of bivalent ligands has facilitated the observed crystal packing. Multivalent ligands are suggested as tools for overcoming difficult crystallisation issues.

## Table of Contents

	<b>Page</b>
Declaration	2
Abstract	3
Table of Contents	4
List of Figures	10
List of Tables	16
1. Introduction	18
1.1. Structure of SAP	21
1.2. Physiological role of SAP	24
1.2.1. DNA binding by SAP	24
1.2.2. Role in Innate Immunity	25
1.2.3. Role in Amyloidosis	25
1.3. Amyloidosis	26
1.3.1. Detection of amyloid deposits	27
1.3.2. Identifying SAP's amyloid binding motif	28
1.3.3. Removing visceral amyloid deposits by targeting SAP	29
1.4. Ligand binding by SAP	29
1.4.1. Small molecule binding	29
1.4.2. Development of bivalent ligands for SAP	31
1.5. Multivalent ligand binding	33
1.5.1. Thermodynamics of multivalent binding	35
1.5.2. Kinetics of multivalent binding	37
1.5.3. Examples of multivalent processes	40
1.5.4. Design of multivalent ligands	42
1.5.5. Multivalency for SAP	45
1.6. Project Aims	46
2. X-ray crystal structure of Serum Amyloid P component in complex with N-	47
	4



## Table of Contents

### acetyl D-proline

2.1. Introduction	47
2.2. Methods and Results	49
2.2.1. Crystallisation of SAP with N-acetyl D-proline	49
2.2.2. X-ray diffraction data collection	51
2.2.3. Data quality assessment	56
2.2.4. Data processing	57
2.2.5. Merging and Scaling	61
2.2.6. Molecular replacement and the phase problem	63
2.2.7. Refinement	66
2.3. Discussion	68
2.3.1. Crystal Packing	68
2.3.2. Protein Structure	68
2.3.3. Structure Validation	70
2.3.3.1. Protein Geometry	70
2.3.3.2. Ramachandran Plot	71
2.3.3.3. B-factor analysis	72
2.3.4. Binding site analysis	73
2.3.5. N-acetyl D-proline as a headgroup for multivalent binding	76

### 3. X-ray crystal structure of rat C-reactive Protein in complex with bivalent ligand

BPC-8	79
3.1. Introduction	79
3.2. Materials and Methods	83
3.2.1. Protein purification	83
3.2.2. X-ray crystallography	83
3.3. Results	84
3.3.1. Protein crystallisation	84
3.3.2. Diffraction data collection and processing	84
3.3.3. Molecular replacement	86
3.3.4. Refinement and model building	87
3.4. Discussion	94
3.4.1. Radiation damage	94

## Table of Contents

3.4.2. Structure analysis	97
3.4.3. Surface charge	99
3.4.4. Crystal packing	101
3.4.5. Extended C-terminus	105
3.4.6. Complement activation	105
3.4.7. Multivalent ligands as tools for crystallography	109
4. Synthesis of multivalent ligands for SAP	111
4.1. Introduction	111
4.1.1. The design of multivalent ligands	111
4.1.1.1. Headgroup design	112
4.1.1.2. Linker design	113
4.1.1.3. Scaffold design	114
4.1.2. Solid Phase Peptide Synthesis	118
4.1.2.2. Synthesis of non-standard peptides	121
4.2. Materials and Methods	121
4.2.1. Synthesis of peptides	124
4.2.2. Small scale cleavage of peptides for mass spectrum monitoring during synthesis	124
4.2.3. Synthesis of central scaffold 5	125
4.2.3.1. SPPS of uncyclised central scaffold 1	125
4.2.3.2. Formation of 2 by addition of chloroacetyl group to the N-terminus of 1	126
4.2.3.3. Cleavage of 2 from the solid phase and purification of product 3	127
4.2.3.4. Removal of TFA from crude 3	128
4.2.3.5. Formation of 4 thioether alkylation	128
4.2.3.6. Formation of 5 by reaction of chloroacetic anhydride with the lysine of 4	129
4.2.4. SPPS of 6, 7 and 8	129
4.2.4.1. Coupling of first amino acid to 2-chlorotrityl chloride resin	129
4.2.4.2. Separation of peptides containing polyethylene glycol from TFA	130
4.2.4.3. Purification of 6, 7 and 8	130
4.2.5. Synthesis of 18, 19 and 20	131

## Table of Contents

4.2.5.1. SPPS of 9, 10 and 11	131
4.2.5.2. Cleaving protected peptides 9, 10 and 11 from 2-Ctr resin using 1% TFA	132
4.2.5.3. Synthesis of 14	132
4.2.5.4. Fragment condensation between decameric linkers (9, 10 and 11) and 14	133
4.2.6. Formation of multivalent ligands	134
4.2.6.1. Preparation of 21, 22, 23, 24, 25 and 26 by thioether formation between 5 and linker-headgroups 6, 7, 8, 18, 19 and 20	134
4.2.6.2. Purification of multivalent ligands and ligand-headgroups	136
4.2.6.3. Preparation of 21-26 as chloride salts	136
4.2.6.4. Packaging and storage of multivalent ligands	137
4.2.7. Ellman test for free thiol groups	137
4.3. Results	138
4.3.1. Final product yields	138
4.3.2. Purity of 21-26	138
4.3.3. Central scaffold cyclisation of thioether formation	139
4.3.4. Multivalent ligand synthesis by thioether formation	140
4.4. Discussion	145
4.4.1. Forming thioether bonds during freeze-drying	145
4.4.2. Identification of multivalent ligands by mass spectrometry	146
4.4.3. Avoiding diketopiperazine formation	147
5. Analysis of binding between SAP and multivalent ligands	148
5.1. Introduction	148
5.2. Materials and Methods	148
5.2.1. Peptide content analysis	148
5.2.2. SAP ligand precipitation assay using Ultraviolet-Visible Spectroscopy	150
5.2.3. Mass Spectrometry of SAP-multivalent ligand mixture	151
5.2.3.1. MS Sample preparation	151
5.2.4. Isothermal Titration Calorimetry of SAP with multivalent ligands	151
5.2.4.1. ITC Sample preparation	155

## Table of Contents

5.2.4.2. ITC of SAP with multivalent ligands 22-26	156
5.2.4.3. ITC of SAP with 21	157
5.2.5. X-ray crystallography SAP with multivalent ligands 23 and 24	157
5.2.5.1. Crystallisation	157
5.2.5.2. Data collection	158
5.2.5.3. Data processing, molecular replacement and structure refinement	158
5.3. Results	158
5.3.1. Peptide content analysis	158
5.3.2. UV-Vis SAP precipitation assay	159
5.3.3. Mass spectrometry of SAP with decavalent ligands	161
5.3.4. ITC of SAP with multivalent ligands	163
5.3.5. X-ray crystallography of SAP in complex with multivalent ligands	166
5.3.5.1. X-ray crystallography of SAP-23	166
5.3.5.2. X-ray crystallography of SAP-24	168
5.4. Discussion	172
5.4.1. Amino acid analysis	172
5.4.2. Ultraviolet/Visible absorption SAP precipitation assay	172
5.4.3. Mass Spectrometry of SAP-multivalent ligand complexes	172
5.4.3.1. Optimising Mass Spectrometry for SAP-decavalent ligand complexes	175
5.4.3.2. Mass Spectrometry of pentavalent ligands 21-23 in complex with SAP	176
5.4.4. Isothermal Titration Calorimetry	176
5.4.4.1. ITC of ligands 22-26	177
5.4.4.2. Stoichiometry deviation from expected values	177
5.4.4.3. Binding cooperativity	178
5.4.4.4. Accurate determination of the dissociation constant	181
5.4.5. X-ray crystallography	183
5.4.5.2. SAP-24 complex structure analysis	183
5.4.6. Cooperative binding of multivalent ligands with SAP	192

## Table of Contents

6. Conclusions	193
7. Acknowledgements	198
8. References	199
A. Appendix	217
A.1. Liquid Chromatography Mass Spectrometry analysis of purified multivalent ligands 21-26	217
A.1.1. Compound 21	217
A.1.2. Compound 22	219
A.1.3. Compound 23	221
A.1.4. Compound 24	223
A.1.5. Compound 25	225
A.1.6. Compound 26	227
A.2. Mass Spectrometry of purified multivalent ligands	229
A.3. Mass Spectrometry of purified central scaffold prior to cyclisation (compound 3) and after cyclisation by intramolecular thioether substitution (compound 4)	232
A.4. Derivation of the dissociation constant, $K_d$ , when half macromolecule, M, binding sites are occupied by ligand, L	233
A.5. Isothermal Titration Calorimetry results for all runs of multivalent ligands binding with SAP	234
A.6. Isothermal Titration Calorimetry binding curves for multivalent ligands 21-26 with SAP	235

## List of Figures

	Page
Figure 1.1. Cartoon representations of the A-face and B-face of a single SAP pentamer.	21
Figure 1.2. Cartoon representation of the structural features of single protomer of SAP.	22
Figure 1.3. Structure of the biantennary oligosaccharide of SAP.	23
Figure 1.4. Cartoon representation of a fibril of Amyloid $\beta$ 1-42 derived from solid state NMR. Computer simulated and cryoelectron micrography images of Amyloid $\beta$ 1-42 fibrils	26
Figure 1.5. Posterior whole body scintigraph after injection of I123 radiolabelled SAP.	28
Figure 1.6. Cartoon and stick representation of phosphothreonine binding to SAP.	30
Figure 1.7. SAP binding ligands identified by Pepys <i>et al.</i> (2002).	31
Figure 1.8. Cartoon representation of x-ray diffraction crystal structures of non-covalent cross-linking of two SAP pentamers by bivalent ligands.	32
Figure 1.9. Mechanisms of monovalent and multivalent binding.	33
Figure 1.10. Schematic of effective concentration of a partially bound multivalent ligand.	36
Figure 1.11. Intramolecular and intermolecular multivalent binding.	38
Figure 1.12. Cartoon representation of a subunit of Concanavalin A.	40
Figure 1.13. a) The STARFISH Shiga-like toxin inhibitor molecule developed by Kitov <i>et al.</i> , 2000. b) STARFISH bound to the B5 subunit of Shiga-toxin 1 (SLT1).	42
Figure 1.14. Examples of rigid central scaffolds utilised for multivalent ligands.	44
Figure 2.1. a) Bis D-proline compound, CPHPC. b) N-acetyl D-proline, NADPro.	48
Figure 2.2. Diagram of hanging drop vapour diffusion crystallisation experiment.	50
Figure 2.3. Protein solubility diagram.	50
Figure 2.4. Diagram of Diamond Light Source Synchrotron (Didcot, UK).	53
Figure 2.5. Diagram of a typical synchrotron beamline.	53
Figure 2.6. Charge-coupled device x-ray detector.	54
Figure 2.7. Bragg's Law and representation of incoming x-ray being diffracted by two parallel planes of atoms within a crystal.	54
Figure 2.8. Diffraction image of NADPro-SAP crystal NAD3.	56
	10

## List of Figures

Figure 2.9. Precession image of NADPro-SAP diffraction intensities of the hk0 plane.	60
Figure 2.10. Enhanced precession image viewing the 0k0 plane.	60
Figure 2.11. Crystal packing of SAP in crystal NAD3.	68
Figure 2.12. A and B faces of NADPro-SAP structure in crystal NAD3.	69
Figure 2.13. Example of 2Fo-Fc electron density map (grey mesh, $1.8\sigma$ ) generated from the NADPro-SAP dataset and the protein model built within it	69
Figure 2.14. Backbone torsion angles of a polypeptide chain.	71
Figure 2.15. Ramachandran plot for the general case of NADPro-SAP protein backbone torsion angles.	72
Figure 2.16. B-factor per residue of each chain for NADPro-SAP structure.	73
Figure 2.17. Ball and stick representation of NADPro bound into the double calcium binding site of SAP.	74
Figure 2.18. Omit map of N-acetyl D-proline electron density calculated at sigma values between 2.5-4.0.	75
Figure 2.19. Calcium-dependent binding of D-proline and L-proline	78
Figure 3.1. The structure of phosphoethanolamine and phosphocholine	79
Figure 3.2. Crystal of bovine enterovirus during room temperature x-ray exposure	81
Figure 3.3. bis-Phosphocholine ligand for CRP, BPC8.	83
Figure 3.4. X-ray diffraction pattern of rCRP-BPC8 crystal during second data collection pass.	85
Figure 3.5. Stick representation of residues 42-46 of rCRP and the corresponding residues of hCRP.	89
Figure 3.6. Cartoon depiction of the x-ray crystal structure of rCRP-BPC8 complex.	90
Figure 3.7. 2Fo-Fc electron density map of residues 83-88 of chain B in rCRP-BPC8 generated in Pymol and the Ramachandran plot for the general case of residues in the rCRP-BPC8 structure.	93
Figure 3.8. Graph of merging R-values for each image (batch) for rCRP-BPC8.	94
Figure 3.9. Electron density map of Cys36-Cys95 disulphide bond.	96
Figure 3.10. Single protomer of rCRP and hCRP viewed along the axis of the $\beta$ -jellyroll.	97
Figure 3.11. The calcium binding site of rCRP in complex with BPC8	98
Figure 3.12. Surface charge map of the B-faces of rCRP and hCRP.	99
Figure 3.13. Surface charge map of A-faces of rCRP and hCRP.	100
Figure 3.14. Crystal packing of rCRP decamers.	102

## List of Figures

Figure 3.15. Example of the potential interprotein contacts formed between neighbouring decameric complexes.	103
Figure 3.16. rCRP A-face charge map interaction.	104
Figure 3.17. Complement binding cleft of hCRP and rCRP with residues identified as important for complement binding.	107
Figure 4.1. Cartoon representation of SAP pentamer demonstrating distances between ligand binding sites.	112
Figure 4.2. Modified CPHPC headgroup amino-CPHPC.	113
Figure 4.3. On-resin cyclisation using Allyl protecting group (Alloc) as used by Zhang et al., 2004.	116
Figure 4.3. General reaction scheme of thioether bond formation by thiol alkylation with an alkyl halide performed in aqueous solvent under mildly basic conditions.	117
Figure 4.4. Chemical structure of each component for the proposed multivalent ligands for SAP and the combined structure	118
Figure 4.5. Generic structure of an $\alpha$ -amino acid with a primary amine.	119
Figure 4.6. Example scheme of the standard reaction steps of Fmoc solid phase peptide synthesis	120
Figure 4.7. Diagram of the sample nebuliser needle and mass analyser of a typical electrospray mass spectrometer (ESI-MS).	123
Figure 4.8. Diagram of the Quadrupole and the path of ions through the mass analyser.	124
Figure 4.9. Synthesis of central cyclic scaffold 5.	126
Figure 4.10. Reaction of ninhydrin with N-terminus of a nascent peptide chain.	127
Figure 4.11. Structure of linker-headgroups 6, 7, 8, 18, 19 and 20.	130
Figure 4.12. Synthesis of decameric linker headgroups 18, 19 and 20.	131
Figure 4.13. Decameric headgroup synthesis.	133
Figure 4.14. Multivalent ligand assembly, thioether formation reaction between central scaffold 5 and linker headgroups 6, 7, 8, 18, 19 and 20.	135
Figure 4.15. Reaction to test for free thiol groups using Ellman's reagent.	137
Figure 4.16. Diketopiperazine formation between glycine and proline during SPPS under basic conditions.	139
Figure 4.17. Mass Spectrum of purified linker-headgroup 6 prior to reaction with central scaffold 5.	140
Figure 4.18. Mass Spectrum of reaction between 5 and 6 in air after 2 hours.	141
Figure 4.19. Mass spectrum of reaction of 5 with 6 in air after 7 hours.	141
Figure 4.20. Maximum entropy mass spectrum of reaction between 5 and 7 in air after drying on freeze-dryer.	144



## List of Figures

Figure 4.21. Maximum entropy mass spectrum of reaction between partially formed 22 and 7 after second freeze-drying of reactants.	144
Figure 5.1. Typical design of calorimeter used in isothermal titration calorimetry	152
Figure 5.2. Typical binding isotherm obtained from an ITC binding experiment.	153
Figure 5.3. Influence of C value on shape of sigmoidal curve.	154
Figure 5.4. Whole Protein Mass Spectrometry results of SAP in complex with multivalent ligands a) 24, b) 25, c) 26 and d) SAP multivalent ligand free.	162
Figure 5.5. Typical results of ITC binding experiment between SAP and multivalent ligand 25.	165
Figure 5.6. Model of the metal-binding site of SAP-23 constructed from the SAP-23 crystal data.	167
Figure 5.7. Left: Fo-Fc difference electron density map drawn at 1.5 $\sigma$ before and after model building.	170
Figure 5.8. Cartoon representation for the SAP-24 crystal structure viewed top down and side on generated with Pymol.	184
Figure 5.9. Structure of ligand 24.	185
Figure 5.10. Ball and stick model of STARFISH Shiga-like toxin inhibitor binding to SLT-1.	186
Figure 5.11. View of the A-face of a single SAP pentamer in the crystal structure of SAP-24 with a proposed position for the central core and polyethylene glycol linkers of ligand 24.	187
Figure 5.12. Calcium-binding site of SAP in SAP-24 crystal structure.	188
Figure 5.13. Comparison of D-proline position for N-CPHPC and NADPro when bound in the calcium-binding site of SAP.	189
Figure 5.14. The electron density map shows a number of cadmium ions coordinated by multiple water molecules and the carboxyl side chain of Glu14.	190
Figure 5.15. Roll of pentamers in SAP-24 and SAP-CPHPC.	191
Figure 6.1. The proposed structure of linker-headgroups for synthesising pentavalent ligands not susceptible to diketopiperazine formation.	194
Figure 6.2. Bivalent ligands for SAP.	195
Figure A.1. Structure of compound 21	217
Figure A.2. UV-Visible spectrum after liquid chromatography of compound 21 at 280nm.	217
Figure A.3. Mass spectrometry detector total ion count recorded during LCMS of purified 21.	218

## List of Figures

Figure A.4. Mass spectrum of peak detected at 5.18 mins.	218
Figure A.5. Structure of compound 22	219
Figure A.6. UV-Vis spectrum after liquid chromatography of compound 22.	219
Figure A.7. Mass spectrometry detector total ion count recorded during LCMS of purified 22.	220
Figure A.8. Mass spectrum of peak detected at 5.43 mins.	220
Figure A.9 Structure of compound 23	221
Figure A.10. UV-Vis spectrum after liquid chromatography of compound 23.	221
Figure A.11. Mass spectrometry detector total ion count recorded during LCMS of purified 23.	222
Figure A.12. Mass spectrum of peak detected at 5.62 mins.	222
Figure A.13. Structure of compound 24	223
Figure A.14. UV-Vis spectrum after liquid chromatography of compound 24.	223
Figure A.15. Mass spectrometry detector total ion count recorded during LCMS of purified 24.	224
Figure A.16. Mass spectrum of peak detected at 5.34 mins.	224
Figure A.17. Structure of compound 25	225
Figure A.18. UV-Vis spectrum after liquid chromatography of compound 25.	225
Figure A.19. Mass spectrometry detector total ion count recorded during LCMS of purified 25.	226
Figure A.20. Mass spectrum of peak detected at 5.80 mins.	226
Figure A.21. Structure of compound 26	227
Figure A.22. UV-Vis spectrum after liquid chromatography of compound 26.	227
Figure A.23. Mass spectrometry detector total ion count recorded during LCMS of purified 26.	228
Figure A.24. Mass spectrum of peak detected at 5.91 mins.	228
Figure A.25. Mass spectrum of compound 21.	229
Figure A.26. Mass spectrum of compound 22.	229
Figure A.27. Mass spectrum of compound 23.	230
Figure A.28. Mass spectrum of compound 24.	230
Figure A.29. Mass spectrum of compound 25.	231
Figure A.30. Mass spectrum of compound 26.	231
Figure A.31. Uncyclised compound 3 prior to intramolecular thioether substitution reaction.	232
Figure A.32. Cyclic compound 4 after intramolecular thioether substitution reaction and purification.	232
Figure A.33. ITC binding curve recorded for ligand 22 with SAP.	235

## List of Figures

Figure A.34. ITC binding curve recorded for ligand 23 with SAP.	235
Figure A.35. ITC binding curve recorded for ligand 24 with SAP.	236
Figure A.36. ITC binding curve recorded for ligand 25 with SAP.	236
Figure A.37. ITC binding curve recorded for ligand 26 with SAP.	237

## List of Tables

	Page
Table 1.1. Sequence alignment for SAP and CRP from humans, the rat and mouse.	18
Table 2.1. Mosflm unit cell determination results for SAP-NADPro crystal prior to full data collection.	58
Table 2.2. Processing and refinement statistics for SAP-NADPro crystal, NAD3.	63
Table 3.1. Aligned amino acid sequences of human CRP and rat CRP.	87
Table 3.2. Processing and refinement statistics for rCRP-BPC8 crystal x-ray diffraction data.	91
Table 3.3. Results of Molprobit analysis of final rCRP-BPC8 model.	92
Table 4.1. Yield of multivalent ligands calculated by weighing mass of chloride salt of 21-26 after freeze-drying.	138
Table 4.2. Purity of multivalent ligands indicated by LCMS.	138
Table 4.3. Predicted and observed masses for multivalent ligands <b>21-26</b> .	147
Table 5.1. Peptide content of multivalent ligand samples as determined by amino acid analysis.	159
Table 5.2. UV-Vis SAP precipitation assay with multivalent ligands 5 minutes after sample preparation.	160
Table 5.3. UV-Vis SAP precipitation assay with multivalent ligands 35 minutes after sample preparation.	161
Table 5.4. Comparison of predicted masses for SAP-decameric ligand complexes versus masses of complexes observed in this study.	163
Table 5.5. ITC results of multivalent ligands 22-26 binding to SAP.	166
Table 5.6. Mosflm unit cell determination results for SAP-24 and unit cell dimensions of PDB file 4AVV.	168
Table 5.7. Processing and refinement statistics for SAP-24 crystal structure.	171
Table 5.8. Deviation of masses of SAP-decavalent ligand complexes from the expected values.	173
Table 5.9. The $\alpha$ -values for the binding constants of the multivalent ligands binding to SAP as measured by ITC.	180
Table A.1. Purity calculation from peak area of LC trace for purified compound 21.	217
	16

## List of Tables

Table A.2. Purity calculation from peak area of LC trace for purified compound 22.	219
Table A.3. Purity calculation from peak area of LC trace for purified compound 23.	221
Table A.4. Purity calculation from peak area of LC trace for purified compound 24.	223
Table A.5. Purity calculation from peak area of LC trace for purified compound 25.	225
Table A.6. Purity calculation from peak area of LC trace for purified compound 26.	227
Table A.7. Isothermal titration calorimetry results for all runs of multivalent ligands 22-26 binding with SAP in the presence of calcium ions.	234

## Chapter 1. Introduction

Serum Amyloid P component, SAP, is a normal circulating plasma glycoprotein in humans (Emsley *et al.*, 1994). As a member of the pentraxin superfamily of proteins (Pepys and Baltz, 1983), it has a distinctive structure of five identical protomers arranged in a pentameric ring. The name pentraxin is derived from the Greek *penta* (five) and *ragos* (berries; Osmand *et al.*, 1977). In addition to the radial protomer arrangement, the pentraxins are related by a 'family signature' amino acid sequence of HxCxS/TWxS (where x is any amino acid; Pepys *et al.*, 1997).

The pentraxin family is divided into the classical 'short' members and the more recently discovered 'long' members. The 'long' pentraxins share a similar C-terminal domain to the 'short' pentraxins, but have an unrelated N-terminal domain (Mantovani *et al.*, 2003). The first 'long' pentraxin discovered was TNF-stimulated gene 14, PTX3 (Lee *et al.*, 1993). Other 'long' pentraxins consist of neuronal proteins NPI, NPII and NPR (Dodds *et al.* 1997) and sperm-egg fusion protein, Apexin (Reid and Blobel, 1994).

The 'short' pentraxins have two known members SAP and C-Reactive Protein, CRP. CRP and SAP share 51% strict sequence identity in humans and 66% homology when conserved residues are taken into account (Pepys *et al.*, 1997). Both 'short' pentraxins are phylogenically conserved plasma proteins found in most vertebrates and some invertebrates with no reported occurrences of deficiency or polymorphism in organisms in which it has been found. However, despite stable conservation through evolutionary history, the exact roles of SAP and CRP vary between species and in many cases remain unclear. The 'short' pentraxins have a variety of roles in different species such as behaviour as an acute phase protein, specificity of ligand binding and ability to activate complement.

## Chapter 1. Introduction

SAP	10	20	30	40	50	60
Homo Sapiens	HTDLSGKVFV	FPRESVTDHV	NLITPLEKPL	Q <sup>N</sup> FTLCFRAY	SDLSRAYS <sup>L</sup> F	SYNTQGRD <sup>N</sup> E
Rattus Norveigicus	QTDLNQKVFV	FPRESETDYV	KLIPWLEKPL	Q <sup>N</sup> FTLCFRAY	SDLSRSQS <sup>L</sup> F	SYSVNSRD <sup>N</sup> E
Mus Musculus	QTDLKRKVFV	FPRESETDHV	KLIPHLEKPL	Q <sup>N</sup> FTLCFR <sup>T</sup> Y	SDLSRSQS <sup>L</sup> F	SYSVKGRD <sup>N</sup> E
	70	80	90	100	110	120
Homo Sapiens	L <sup>L</sup> VY <sup>K</sup> ERVGE	YSL <sup>Y</sup> IGRH <sup>K</sup> V	TSKVIEK <sup>F</sup> PA	PVHICVSW <sup>E</sup> S	SSGIAEFW <sup>I</sup> N	GTPLVKK <sup>G</sup> LR
Rattus Norveigicus	L <sup>L</sup> IY <sup>K</sup> AKLEQ	YSL <sup>Y</sup> IGNS <sup>K</sup> V	TVRGLEEF <sup>P</sup> S	PIHFCTSW <sup>E</sup> S	SSGIAEFW <sup>I</sup> N	GKPWVK <sup>G</sup> KLQ
Mus Musculus	L <sup>L</sup> IY <sup>K</sup> EKVG <sup>E</sup>	YSL <sup>Y</sup> IGQSK <sup>V</sup>	TVRGMEE <sup>Y</sup> LS	PVHLCTTW <sup>E</sup> S	SSGIVEFW <sup>I</sup> N	GKPWVK <sup>G</sup> SLQ
	130	140	150	160	170	180
Homo Sapiens	QGYFVEAQ <sup>P</sup> K	IVLGQEQD <sup>S</sup> Y	GGKFDRSQ <sup>S</sup> F	VGEIGDLY <sup>M</sup> W	DSVLPPEN <sup>I</sup> L	SAYQGTPL <sup>P</sup> A
Rattus Norveigicus	KGYTVKSS <sup>P</sup> S	IVLGQEQD <sup>T</sup> Y	GGGFDKTQ <sup>S</sup> F	VGEIADLY <sup>M</sup> W	DSVLTPEN <sup>I</sup> H	SVDRGFPP <sup>N</sup> P
Mus Musculus	REYTVKAPP <sup>S</sup>	IVLGQEQD <sup>N</sup> Y	GGGFQRSQ <sup>S</sup> F	VGEFSDLY <sup>M</sup> W	DYVLTPQD <sup>I</sup> L	FVYRDSPV <sup>N</sup> P
	190	200	210	220		
Homo Sapiens	NILDWQAL <sup>N</sup> Y	EIRGYV <sup>I</sup> IKP	LVWV			
Rattus Norveigicus	NILDWRAL <sup>N</sup> Y	EINGYV <sup>V</sup> IKP	RMWDNK <sup>S</sup> S			
Mus Musculus	NILNWQAL <sup>N</sup> Y	EINGYV <sup>V</sup> IRP	RVWD			
CRP	10	20	30	40	50	60
Homo Sapiens	QTDMSRKAFV	FPKESDTSYV	SLKAPLT <sup>K</sup> PL	KAFTVCL <sup>H</sup> FY	TELSSTRG <sup>Y</sup> S	IFS <sup>Y</sup> ATKRQD <sup>G</sup>
Rattus Norveigicus	HEDMSKQAFV	FPGVSATAYV	SLEAESK <sup>K</sup> PL	EAFTVCLY <sup>A</sup> H	ADVS--RS <sup>F</sup> S	IFS <sup>Y</sup> ATKTS <sup>F</sup>
Mus Musculus	HEDMFKKAFV	FPKESDTSYV	SLEAESK <sup>K</sup> PL	NTFTVCL <sup>H</sup> FY	TALSTVRS <sup>F</sup> S	VFS <sup>Y</sup> ATKK <sup>N</sup> S
	70	80	90	100	110	120
Homo Sapiens	NEILL <sup>F</sup> W <sup>S</sup> KD	IGYSFTVG <sup>G</sup> S	E <sup>I</sup> ILFEVPE <sup>V</sup> T	VAPVHICT <sup>S</sup> W	ESASGIVE <sup>F</sup> W	VDGKPRV <sup>R</sup> KS
Rattus Norveigicus	NEILL <sup>F</sup> W <sup>T</sup> RG	QGFSIAVG <sup>G</sup> P	E <sup>I</sup> ILFSASE <sup>I</sup> P	EVPTHICAT <sup>W</sup>	ESATGIVEL <sup>W</sup>	LDGKPRV <sup>R</sup> KS
Mus Musculus	NDILL <sup>F</sup> W <sup>N</sup> KD	KQYTFGVG <sup>G</sup> A	E <sup>V</sup> RFMVSE <sup>I</sup> P	EAPTHICAS <sup>W</sup>	ESATGIVE <sup>F</sup> W	IDGKPKV <sup>R</sup> KS
	130	140	150	160	170	180
Homo Sapiens	LKKGYTVGAE	ASIILGQEQD <sup>G</sup>	SFGGNFEG <sup>S</sup> Q	SLVGDIGN <sup>V</sup> N	MWDFVLSP <sup>D</sup> E	INTIYLG <sup>G</sup> PF
Rattus Norveigicus	LQKGYIVGT <sup>Q</sup>	ASIILGQEQD <sup>G</sup>	SYGGGFDA <sup>N</sup> Q	SLVGDIGD <sup>V</sup> N	MWDFVLSP <sup>D</sup> Q	INAVYVGR <sup>V</sup> F
Mus Musculus	LHKG <sup>Y</sup> TVGPD	ASIILGQEQD <sup>G</sup>	SYGGGFDA <sup>K</sup> Q	SLVGDIGD <sup>V</sup> N	MWDFVLSPE <sup>Q</sup>	ISTVYVGG <sup>T</sup> L
	190	200	210	220		
Homo Sapiens	SPNVLNWRAL	KYEVQGEV <sup>F</sup> T	KPQLW <sup>P</sup>			
Rattus Norveigicus	SPNVLNWRAL	KYETHGDV <sup>F</sup> I	KPQLWPL <sup>T</sup> DC	CES		
Mus Musculus	SPNVLNWRAL	NYKAQGDV <sup>F</sup> I	KPQLW <sup>S</sup>			

Table 1.1. Sequence alignment for SAP and CRP from humans, the rat and mouse. Residues involved in metal binding are highlighted (turquoise) along with residues adjacent to the metal binding site, which are often involved in ligand binding (red) such as the hydrophobic pocket formed by Leu62, Tyr64 and Tyr74 in SAP and glycosylation sites (yellow).

CRP is an acute phase protein in humans. An acute phase response is a physiological and biochemical response to forms of tissue injury, inflammation or infection (Pepys and Hirschfield, 2003). There are a range of acute phase proteins some of which increase in concentration, others decrease in concentration in response to stimuli. For CRP, the acute phase response can be dramatic, with the concentration increasing as much as 1000-fold from a normal plasma concentrations of less than 10mg L<sup>-1</sup> (Gabay

and Kushner, 1999; Young *et al.*, 1991). CRP is also known to activate complement by binding to the C1q recognition protein (Kaplan and Volanakis, 1974; Siegel *et al.*, 1974). CRP-mediated complement activation has subsequently been implicated as the cause of increased inflammation and tissue injury during myocardial infarction and stroke (Griselli *et al.*, 1999; Gill *et al.*, 2004).

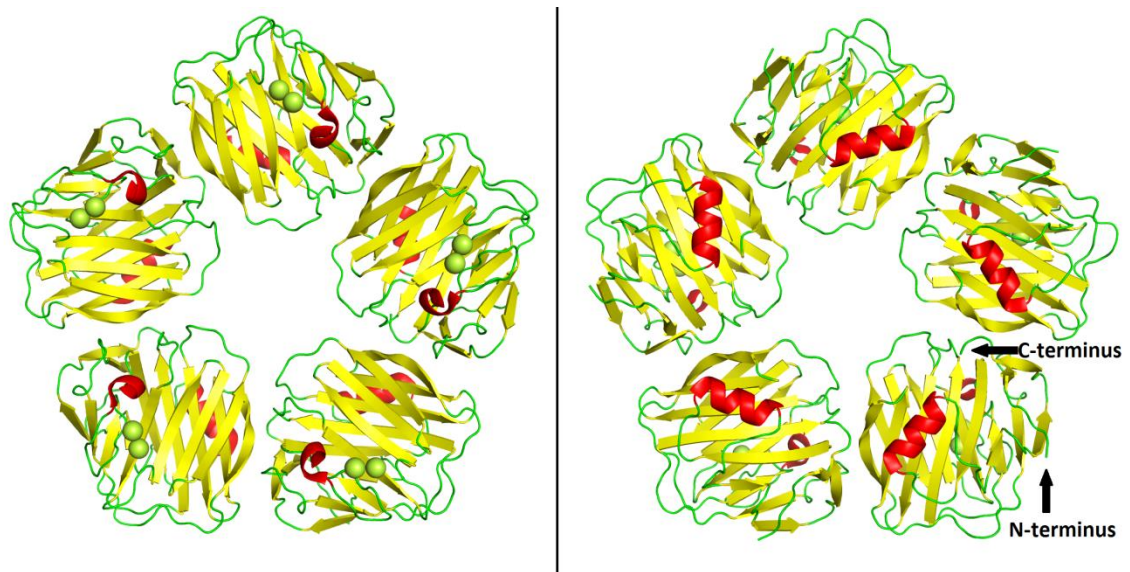
SAP, as its name suggests, is associated with universal presence in amyloid deposits (Bladen *et al.*, 1966; Cathcart *et al.*, 1967). The amyloidoses are a heterogeneous group of disorders characterised by extracellular deposition of abnormally folded proteins (Tan and Pepys, 1994). A range of proteins and peptides are known to form amyloid fibres and the process of asymptomatic amyloid deposition is a consequence of ageing. Clinical amyloidosis occurs in a range of diseases, for example the formation of  $\beta$ -amyloid plaques occurs dramatically in Alzheimer's disease (Hardy and Higgins, 1992),  $\beta$ -2-Microglobulin amyloid deposition is a complication of long-term haemodialysis (Gorevic *et al.*, 1986) and formation of islet amyloid is a common feature in patients with Type II Diabetes (Jiménez *et al.*, 2002). SAP is universally present in all amyloid deposits regardless of the nature of the amyloid protein and constitutes around 15% of the dry mass. SAP in amyloid deposits coats and stabilises fibrils against attack from phagocytes (Tennent *et al.*, 1995). It is likely that SAP is binding to a common motif present on all amyloid proteins, but as yet no motif has been identified.

For the aforementioned reasons, CRP and SAP have emerged as targets for therapeutics in the treatment of cardiovascular disease and amyloidosis respectively (Pepys *et al.*, 2006; Kolstoe and Wood, 2010). Drug discovery programs are currently ongoing for both CRP and SAP. This section will focus principally on the SAP, with information regarding CRP introduced in Chapter 3 where appropriate. Firstly, the structure of SAP will be introduced followed by a discussion of the possible physiological role of SAP and its role in amyloid deposits. The Chapter will also review ligand binding by SAP, development of ligands for SAP and conclude with an introduction to multivalent binding processes and emerging knowledge of their role in biological processes, as well as some potential applications in targeted therapeutics.



### 1.1. Structure of SAP

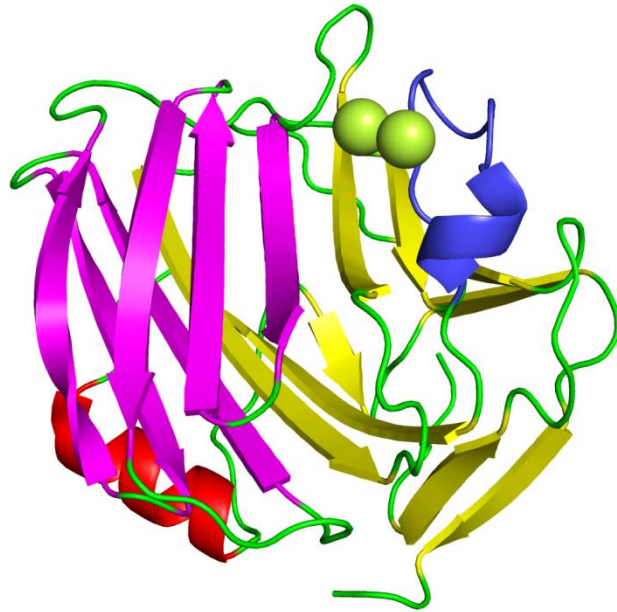
SAP and CRP have 51% strict sequence identity in man and have very similar tertiary and quaternary structures (Ashton *et al.*, 1997). The short Pentraxins are recognisable by their homopentameric quaternary structure (Osmand *et al.*, 1977). Their structure consists of five identical protomers arranged in a pentameric disc-like configuration (Figure 1.1). Pentameric SAP measures approximately 100Å in diameter, 35Å deep with a central pore 20Å in diameter.



**Figure 1.1.** Cartoon representations of the B-face (Left) and A-face (Right) of a single SAP pentamer. Drawn using molecular graphics program Pymol (Pymol molecular graphics system, version 1.3. Distributed by Schrödinger LLC) using SAP model coordinates 1SAC downloaded from the protein data bank (PDB; [www.rcsb.org/pdb](http://www.rcsb.org/pdb)). The model shows five identical protomers formed from  $\beta$ -sheets (yellow) linked by short loop regions (green) arranged in a  $\beta$ -jellyroll topology, the short  $\alpha$ -helix (red) on the “A-face” of the pentamer and two calcium ions (limon yellow) on the opposing binding, “B-face”.

Each protomer of SAP consists of 204 amino acids folded into a flattened  $\beta$ -jellyroll topology formed by two anti-parallel  $\beta$ -sheets linked by short loop regions (Figure 1.2; Emsley *et al.*, 1994). The core secondary structure of SAP is remarkably similar to that of legume lectins such as Concanavalin A. Both  $\beta$ -sheets are extended by a  $\beta$ -meander of three further strands. The hydrophobic core of the protomer forms non-polar interactions with neighbouring protomers, burying around 15% of the total surface area and resulting in a very strong attraction that requires boiling in sodium dodecyl sulphate to dissociate the protomers (Pepys *et al.* 1997). The  $\beta$ -jellyroll is stabilised by multiple hydrogen bonds and a single intramolecular disulphide bond between Cys36-Cys95.

The loop regions which link the secondary structure elements are short, which gives the protomers a tight compact appearance and makes SAP highly resistant to proteolysis.

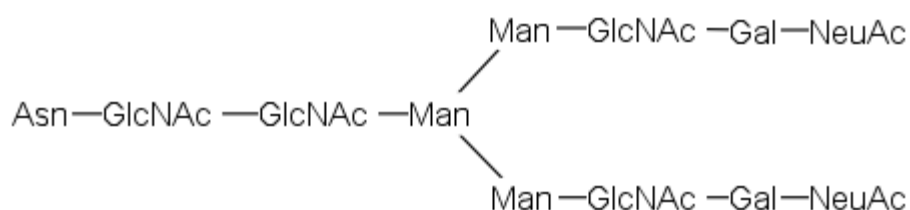


**Figure 1.2.** Cartoon representation of the structural features of single protomer of SAP. The  $\beta$ -jellyroll is depicted by the yellow ribbon with the extended  $\beta$ -meander in purple. A short  $\alpha$ -helix is visible on the A-face of the protomer with two calcium ions (limon yellow) shown bound on the opposing B-face. The calcium binding loop is highlighted in blue and all other loop regions are represented by green. Image generated in Pymol using the structure of SAP was downloaded from the PDB (PDB file: 1SAC).

A short  $\alpha$ -helix is found on one face of each subunit, the so-called 'A-face'. The  $\alpha$ -helix contains Glu167, which is the residue responsible for calcium-dependent aggregation of SAP in the absence of any other ligand. The opposing side of the protomer, commonly termed the 'B-face' contains the double-calcium binding site of SAP. The calcium binding site is formed by residues from a  $\beta$ -sheet and a flexible loop which undergoes a conformational change in the presence of calcium. The loop is composed of residues 134-152 containing mostly acidic residues capable of coordinating the divalent cations. At physiological pH and in the absence of calcium, SAP is detected as a decameric complex in which two SAP pentamers interact through their disordered calcium-binding loops (Pepys *et al.*, 1977). The disordered loop region is the only site at which proteolysis can occur and this is completely inhibited by the binding of calcium ions (Kinoshita *et al.*, 1992).

The calcium binding of SAP coordinates two metal ions separated by around 4Å. The first calcium ion is bound by the side chains of Asp58, Asn59, Glu136 and Asp138 and the main chain carbonyl oxygen of Gln137. The second calcium ion is more loosely bound being that it is coordinated by two water molecules with the side chains of Glu136, Asp138 and Asn148 providing the only protein ligands (Mikolajek *et al.*, 2011). The coordination spheres of both cations are completed by the phosphate or carboxylate group of a ligand. A small hydrophobic pocket adjacent to the calcium binding site formed by the side chains of Leu62, Tyr64 and Tyr74. Though small, the hydrophobic pocket is known to be important for efficient ligand binding to SAP. The side chain pyrrolidine ring of N-acetyl D-proline and a methyl group of 4,6-*O*-(1-carboxyethylidene)- $\beta$ -D-galactosepyranoside (MO $\beta$ DG) are two ligands known to interact with the hydrophobic pocket. There is also some evidence that carbohydrate based ligands, such as MO $\beta$ DG, form electrostatic interactions between oxygen atoms in their heterocyclic ring and the amide groups of Gln148 and Asn59 (Thompson *et al.*, 2002).

A single N-glycosylation site exists on the side chain of Asn32 of each protomer (Tennent and Pepys, 1994). The single biantennary oligosaccharide is located towards the A-face of the protomer at the N-terminal end of a  $\beta$ -sheet that is obscured from solvent by the  $\alpha$ -helix. Unusually, the oligosaccharide does not exhibit the same extent of microheterogeneity common in plasma glycoproteins, particularly during an acute phase response (Figure 1.3). Loss of the sialic acids from the terminus of the oligosaccharide results in glycosylated pentraxins being rapidly catabolised by hepatocytes (Pepys *et al.*, 1994; Yang *et al.*, 1992).



**Figure 1.3.** Structure of the biantennary oligosaccharide of SAP. Image generated using Marvin sketch (Marvin Sketch, version 5.12.0, Chemaxon Ltd.).

## 1.2. Physiological role of SAP

The stable conservation of SAP and the lack of any polymorphism in man, strongly suggests an important physiological function, but despite many theories, its exact role remains unclear. This lack of clarity is due to numerous difficulties in its study, for example the lack of any known deficiency or polymorphism means the results of such an affliction cannot be directly observed. Furthermore, it is difficult to extrapolate results of experiments using animal models because the pentraxins fulfil different roles in different species. For example, murine SAP is an acute phase protein (Pepys *et al.*, 1979), increasing in plasma concentration by up to 50-fold during an acute phase response; in humans, only modest increases of SAP concentration are observed and only during chronic active diseases (Pepys *et al.*, 1997). SAP binds ligands in a calcium-dependent manner, but it is also capable of binding to neighbouring pentamers through non-covalent interaction between the calcium binding site and Glu167. This presents a problem for *in vitro* study because purified SAP rapidly aggregates at physiological calcium concentration and pH and is only inhibited by the presence of a ligand or calcium chelating species such as EDTA (Hind *et al.*, 1984).

### 1.2.1. DNA binding by SAP

One interesting observation is that SAP is the only normal plasma protein capable of binding to DNA and chromatin. The binding displaces H1-type histones, thus solubilising the DNA (Pepys and Butler, 1987; Butler *et al.*, 1990). It could be that SAP is involved in controlling the degradation of extracellular chromatin during cell death (Bickerstaff *et al.*, 1999). This potential role is supported by the tightly folded structure of SAP; secondary structures are linked by short loop regions, preventing proteolysis in the presence of calcium ions, which suggests SAP operates in a hostile environment (Kinoshita *et al.*, 1992). The marshalling of this potentially biologically active cellular debris may be important in preventing autoimmunity and stopping incorrect genetic information entering neighbouring cells prior to phagocytosis (Breathnach *et al.*, 1989). Furthermore, evidence is emerging that SAP's role in mediating DNA clearance could be the reason for the poor response to DNA vaccinations observed in Humans (Wang *et al.*, 2012).

SAP has also been observed binding to the surface of apoptotic cells (Familian *et al.*, 2001). This may be due to chromatin or phosphoethanolamine fragments exposed on the surface of dying cells. It has been suggested that the binding of SAP to apoptotic cells facilitates their uptake by macrophages, which could have implications in the treatment of systemic lupus erythematosus where clearance of apoptotic cells is hampered (Bijl *et al.*, 2003).

### **1.2.2. Role in Innate Immunity**

SAP's role in innate immunity is slightly paradoxical; the rate of phagocytosis is retarded for some bacteria to which SAP binds, but SAP knock out mice are more susceptible to lethal infection with *E. Coli*. O111:B4, which SAP does not bind (Noursadeghi *et al.*, 2000). This could be explained by observations that SAP protects podocytes in the renal system from Haemolytic Uremic Syndrome by binding to Shiga toxin 2 produced by *E. Coli*. (Dettmar *et al.*, 2014). Therefore SAP may have some beneficial host defence function, but it has been suggested that SAP's beneficial function in mediating DNA removal has been hijacked by evolution of some bacteria using bound SAP as camouflage from the immune system.

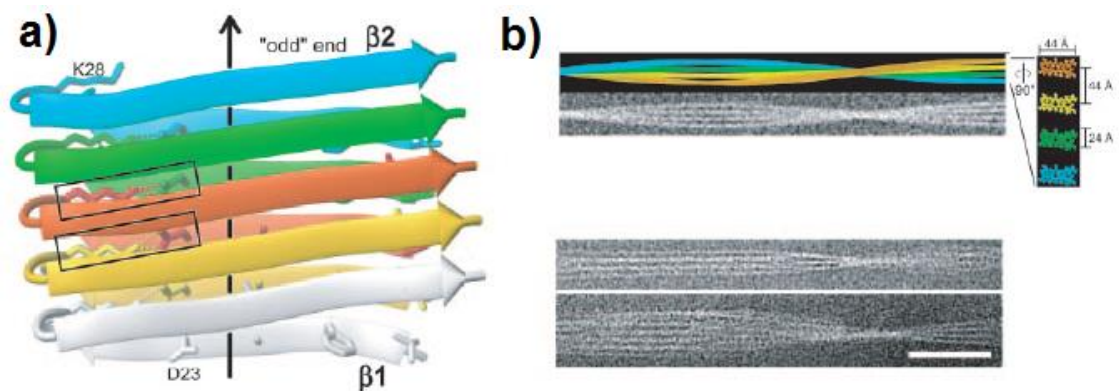
### **1.2.3. Role in Amyloidosis**

Since 1966, SAP has been seen in amyloid deposits as its identical Amyloid P component (AP), although originally mistaken for amyloid fibres (Bladen *et al.*, 1966). In 1967 AP was identified as the key plasma constituent of amyloid, accounting for 15% of the dry mass of deposits (Cathcart *et al.*, 1967; Pepys *et al.*, 1997) and was subsequently isolated in its identical serum form (Pepys *et al.*, 1977). SAP binds to all forms of amyloid fibril in a calcium-dependent manner regardless of the amyloid peptide present (Pepys *et al.*, 1994). Tennent *et al.* (1995) identified that in binding to amyloid deposits, SAP stabilises fibrils, preventing their proteolysis. For this reason, SAP's role in amyloidosis has attracted significant interest as a target for therapeutics to treat amyloidosis and amyloid-related conditions.

### 1.3. Amyloidosis

Amyloid deposits were first discovered in 1851 when Virchow initially incorrectly suggested cellulose as the principal constituent based on results of staining fibrils with iodine (Virchow, 1851). The staining was actually caused by the presence of heparan sulfate proteoglycans in amyloid deposits. Friedrich and Kekule subsequently correctly identified protein to be the major polymeric constituent of amyloid (Friedrich and Kekule, 1859).

Amyloid proteins coalesce to form fibrils, which have a common quaternary structures regardless of the protein involved (Sunde *et al.*, 1997). Amyloid fibrils consist of twisted fibres of  $\beta$ -strands running parallel to the axis of the fibrils, and with individual protein chains perpendicular to the long axis of the fibrils (Figure 1.4). Approximately 50 different disorders associated with the misfolding of normally soluble peptides and proteins (reviewed by Knowles *et al.*, 2014). As the mass of amyloid fibrils grows, they associate strongly with plasma and extracellular matrix proteins and proteoglycans forming amyloid deposits, which invade the extracellular space of organs destroying normal tissue architecture and function (Sipe, 1992). In addition to similar morphology, amyloid fibrils are characterised by resistance to proteolytic digestion (Sorenson and Shimamura, 1964) and insolubility under physiological conditions (Glenner *et al.*, 1968).



**Figure 1.4.** a) Cartoon representation of a fibril of Amyloid  $\beta$  1-42 derived from solid state NMR. b) Top: simulation of fibril of Amyloid  $\beta$  1-42 formed from four protofilaments. Bottom: cryoelectron micrographs of Amyloid  $\beta$  1-42 fibrils. Figure adapted from Lührs *et al.*, 2005.

Amyloidosis is a clinical disorder which results in extracellular deposition of abnormal fibrils derived from aggregation of misfolded, normally soluble protein (Pepys, 2006).

Amyloidosis is directly responsible for 1 in 1000 deaths in developed countries. The clinical disorder can either be systemic or local. Systemic amyloidosis occurs when circulating plasma protein precursors are deposited as amyloid in some or all of the viscera, blood vessel walls and connective tissue (Sipe, 1992). In local amyloidosis, amyloid deposits are restricted to a single organ or tissue.

Amyloidosis may be acquired as a complication of a primary disease that causes occurrence of abnormal protein or increased quantities of a normal protein with amyloidogenic properties, for example amyloid light-chain (AL) amyloidosis is caused by increased production of Ig light chain protein, which commonly accumulates in the Kidneys (Pepys, 2006; Gertz *et al.*, 2002). Alternatively, amyloidosis can be hereditary and caused by gene mutations that encode a variant protein structure more susceptible to amyloidogenesis, such as transthyretin-related hereditary amyloidosis, which is caused by inherited mutations of the gene encoding the protein transthyretin (Saraiva, 1995).

Amyloid deposits may form as a result of other conditions, these are distinct from the clinical disorder of amyloidosis because they are not the direct cause of the clinical condition. For example,  $\beta$ -amyloid plaques form in the brains of patients with Alzheimer's disease, but there is a poor correlation between the amount of amyloid deposited and cognitive impairment. However, the presence of oligomers of  $\beta$ -amyloid, which are a precursor to amyloid fibrils, has been shown to increase cognitive impairment (Cleary *et al.*, 2005). The fact that SAP assists amyloidogenesis (Botto *et al.*, 1997) and stabilises amyloid fibrils has led to suggestions that SAP may assist amyloid fibril formation as a way to remove toxic amyloid oligomers from circulation. This may be supported by observations of SAP's activity as a chaperone made by Coker *et al.* (2000).

### **1.3.1. Detection of amyloid deposits**

The characteristic test for presence of amyloid is staining with Congo Red dye *in vitro* and observing the resulting birefringence (Puchtler *et al.*, 1962). Hawkins and Pepys (1995) have developed a method for identifying amyloid deposits *in vivo* using radiolabelled  $I^{123}$ -SAP. SAP's strong affinity for amyloid means the attached iodine

isotope accumulates in deposits, which are then detected by whole body scintigraphy (Figure 1.5).



**Figure 1.5.** Posterior whole body scintigraph after injection of  $I^{123}$  radiolabelled SAP. Abnormal uptake of the radioactive tracer is observed in the spleen, liver and bone in a patient with systemic Amyloid Light-chain (AL) amyloidosis. The left pane shows amyloid deposition prior to treatment. The right pane shows the same patient 6 months later after 4 rounds of chemotherapy. Adapted from Hawkins and Pepys (1995).

### 1.3.2. Identifying SAP's amyloid binding motif

Despite significant interest, no SAP binding motif has been identified for any amyloid protein. SAP does not bind to the proteins in their native state, indicating that there may be some common motif present in the abnormally folded proteins, which SAP recognises as an amyloid protein (Pepys *et al.*, 1997). It has been suggested that SAP may be binding to the non-protein constituents common to amyloid deposits such glycosaminoglycans, dermatan and heparan sulphate (Hamazaki, 1987; Nelson *et al.*, 1991). However, SAP has been observed binding to isolated amyloid fibrils *in vitro*. Therefore, it is conceivable that SAP binding may involve a mixture of different ligands from the fibrils and proteoglycan components of an amyloid deposit.



### 1.3.3. Removing visceral amyloid deposits by targeting SAP

Amyloid therapies that target SAP have been under development for the past two decades (Pepys *et al.*, 2002; Ho *et al.*, 2005). A promising method for clearing visceral amyloid deposits involves a two-stage treatment in which first circulating SAP is depleted using small molecule bivalent ligand, CPHPC (Gillmore *et al.*, 2010; Figure 1.7). Once circulating SAP has been depleted then an anti-SAP antibody can be administered, which directs to amyloid deposits containing the only significant quantities of remaining SAP (Bodin *et al.*, 2010). This has been shown to be an effective method of highlighting the presence of amyloid to the immune system, which results in rapid clearance of deposits.

## 1.4. Ligand binding by SAP

### 1.4.1. Small molecule binding

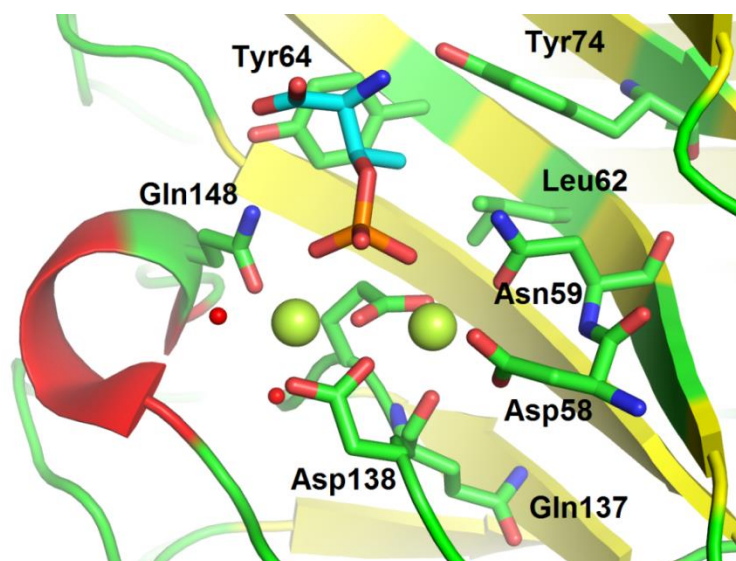
A range of natural and synthetic molecules are known to bind to SAP in a calcium-dependent manner. All known small molecules bind through either a phosphate or carboxylate group coordinating the double-calcium binding site of SAP. Additional protein-ligand interactions also involve, to a greater or lesser extent, the hydrophobic pocket adjacent to the calcium-binding site and electrostatic interactions between the ligand and polar side chain functional groups of the protein.

Phosphoethanolamine (PE) is the archetype synthetic ligand for SAP (Pontet *et al.*, 1978), with phosphocholine (PC) the corresponding ligand for CRP, although CRP also binds to PE (Mikolajek *et al.*, 2011). This binding preference is used in the isolation and purification of the pentraxins from human sera (Hawkins *et al.*, 1991). PE- and PC-sepharose columns are used sequentially to bind the pentraxins with high specificity in the presence of calcium. Prior to the development of immobilised PE columns, SAP was purified using agarose gel (Pepys *et al.*, 1977) with the R-enantiomer of the cyclic pyruvate of galactose, MO $\beta$ DG, identified as the SAP binding species.

MO $\beta$ DG is one of a number of carbohydrate-based ligands for SAP, which also include glycosaminoglycans such as heparin (Thompson and Enfield, 1978), in particular heparan and dermatan sulphates, glycans terminating in mannose and glycans with pre-

terminal galactose (Pepys *et al.*, 1997). Binding is not limited to saccharides; nucleotide deoxyadenosine monophosphate (dAMP) and a range of amino acids, such as D-proline also bind to SAP in a calcium-dependent manner.

SAP binds more strongly to D-amino acids than the analogous L-enantiomer in all reported cases. For example, the N-acetyl D-proline binds to SAP with a 15 $\mu$ M dissociation constant,  $K_d$ , the comparative interaction of N-acetyl L-proline has a  $K_d$  of only 322 $\mu$ M (Kolstoe thesis, 2005). This difference in binding affinity can be explained by the ability of the pyrrolidine ring of NADPro to fit into the hydrophobic pocket of SAP. Other amino acids such as D-valine and D-alanine have been observed binding to SAP through their C-terminus either by isothermal titration calorimetry or x-ray diffraction (Kolstoe thesis, 2005; Purvis thesis, 2002). In contrast, O-phosphothreonine binds through the side chain phosphate in preference to the C-terminal carboxyl group (Figure 1.6). This interaction also serves to underline the importance of the hydrophobic interaction, as the side chain methyl group fits into the hydrophobic pocket.



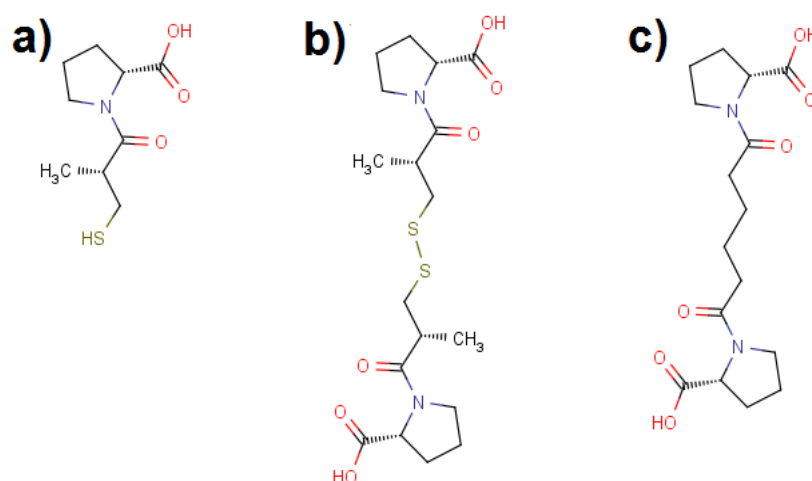
**Figure 1.6.** Cartoon and stick representation of phosphothreonine binding to SAP (Generated with Pymol; PDB file: 2W08). Two calcium ions (limon yellow) are coordinated by multiple side chains of the protein, with the coordination completed by water molecules (red spheres) and the phosphate group of the ligand. The methyl group of the side chain of phosphothreonine can be seen in the hydrophobic pocket formed by Leu62, Tyr64 and Tyr74. Cartoon:  $\alpha$ -helix (red),  $\beta$ -sheet (yellow), loops (green). Protein stick representation: carbon (green), oxygen (red) and nitrogen (blue). Phosphothreonine stick diagram: carbon (turquoise), oxygen (red), nitrogen (blue), phosphorus (orange).

L-glutamic acid binds to SAP through its side chain carboxylic acid, although this has only been observed when Glu167 from the A-face of an SAP pentamer occupies the B-face of another pentamer during auto-aggregation. The auto-aggregation of SAP is not prevented by addition of 100mM free L-glutamic acid. This suggests a multivalent interaction where the presence of multiple Glu167 in a single binding entity presented in a pentagonal symmetry is stronger than the monomeric interaction.

#### 1.4.2. Development of bivalent ligands for SAP

The homopentameric symmetry of SAP suggests simultaneous binding of multiple ligands. This has been surreptitiously exploited by the development of bivalent ligands that non-covalently cross-link SAP forming tightly bound decameric complexes with binding affinities indicative of cooperative binding (Figure 1.8).

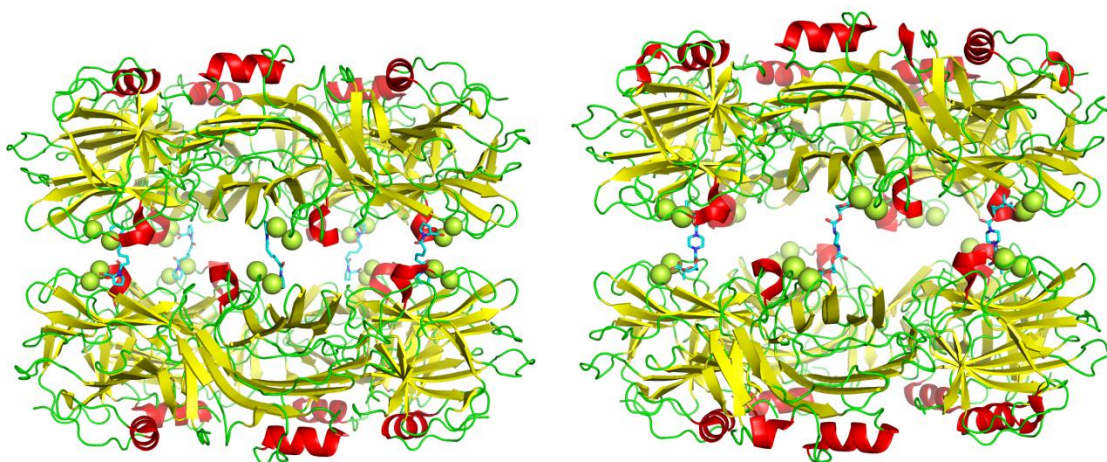
The D-enantiomer of angiotensin converting enzyme inhibitor, Captopril (Smith and Vane, 2003), was an initial hit compound for SAP discovered during a search of the Roche compound library (Pepys *et al.*, 2002). It was observed that dimerising D-Captopril by disulphide bond formation increased the  $IC_{50}$  from 100 $\mu$ M for monomeric D-Captopril to 5 $\mu$ M for the dimeric compound (Figure 1.7). Rationalisation of the structure yielded R-1-[6-[R-2-carboxy-pyrrolidin-1-yl]-6-oxo-hexanoyl]pyrrolidine-2-carboxylic acid (CPHPC). CPHPC is simply two D-proline headgroups linked by a hexanoyl chain, which binds to SAP with a remarkable 10nM dissociation constant.



**Figure 1.7.** SAP binding ligands identified by Pepys *et al.* (2002). a) D-Captopril. b) D-captopril dimer. c) CPHPC.

CPHPC has been shown to have significant potential in the treatment of amyloidosis and amyloid-related conditions. The cross-linking binding of SAP is preserved *in vivo*, leading to rapid catabolism of decameric protein by the hepatocytes and depletion of SAP from circulating plasma (Gillmore *et al.*, 2010). However, despite binding with a low nanomolar binding affinity, CPHPC does not strip SAP from amyloid fibres (Wood and Kolstoe, 2010). This quirk allows SAP in amyloid deposits to be targeted by an anti-SAP antibody after depletion of the protein from plasma. Bodin *et al.* (2010) have shown this methodology to be an effective way of rapidly clearing visceral amyloid deposits *in vivo*.

Ho *et al.* (2005) have also developed bivalent pyruvate acetal of glycerol-based ligands for SAP through rational optimisation of Mo $\beta$ DG's binding to SAP. In both cases, the bivalent ligands induce non-covalent decamer formation in SAP leading to a dramatic increase in binding affinity.



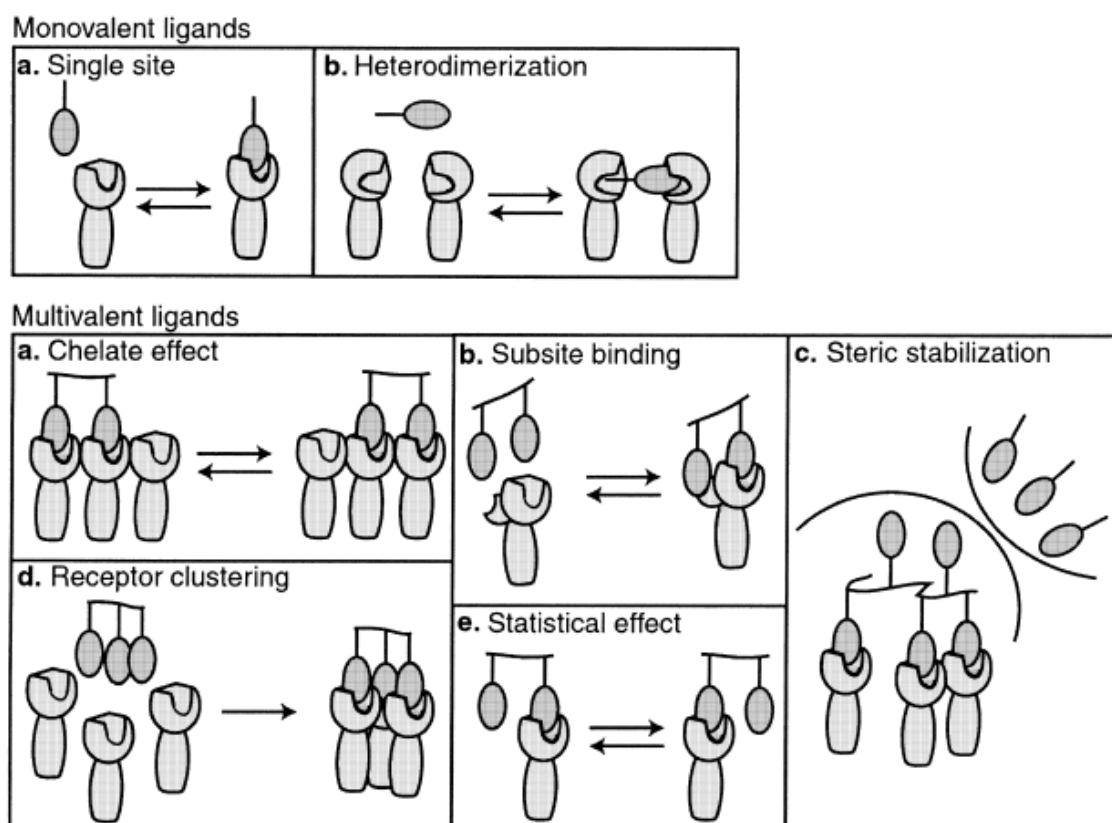
**Figure 1.8.** Cartoon representation of x-ray diffraction crystal structures of non-covalent cross-linking of two SAP pentamers by bivalent ligands (stick representation). Left: CPHPC-SAP decameric crystal structure (PDB file: 4AVT; Kolstoe *et al.*, 2014). Right: Pyruvate acetal of glycerol bivalent ligands in complex with SAP (PDB file: 2A3X; Ho *et al.*, 2005) Model generated in Pymol, cartoon colours:  $\alpha$ -helix (red),  $\beta$ -sheet (yellow), loops (green), calcium ions (limon yellow). Bivalent ligand colours: carbon (turquoise), oxygen (red), nitrogen (blue).

The cooperative mode of binding displayed by bivalent ligands suggests that further binding affinity increases might be achieved by increasing the valency of the binding species further. The natural conclusion based on the pentameric structure of SAP is the synthesis of pentavalent ligands, while the evidence of beneficial pharmacokinetic

properties of cross-linking by bivalent ligands provides the motivation for synthesising decavalent ligands that maintain the non-covalent cross-linking ability of CPHPC. A firm understanding of multivalency is crucial to realising the maximum potential affinity of polyvalent ligands for SAP.

### 1.5. Multivalent ligand binding

A multivalent interaction is defined as the binding of multiple ligands from a single entity, e.g. a surface or molecule, to multiple receptors from another binding entity (Mammen *et al.*, 1998). Presentation of monovalent ligands on an appropriate entity can turn a moderate monovalent binding affinity into a strong multivalent interaction through a range of mechanisms distinct from those of the corresponding monovalent interaction (Figure 1.9; reviewed by Gestwicki *et al.*, 2002). However, the effect of multivalency is not limited to simply an affinity increase, it can also increase the specificity of binding, allowing development of more targeted molecules.



**Figure 1.9.** Mechanisms of monovalent and multivalent binding adapted from Gestwicki *et al.* (2002).

Subsite binding occurs when the interaction of a multivalent ligand with the main binding site facilitates a subsequent interaction at a lesser binding site, which may not

be possible for a monovalent ligand because the ligand binds more loosely to the subsite (Chittasupho, 2012). Chelation effects occur when the simultaneous binding of multiple ligands from the same entity decreases the rate of dissociation. Steric stabilisation is a benefit derived from the larger size and distribution of a multivalent ligand over the monovalent species, which prevents binding of other competitive ligands. Statistical binding occurs when the presentation of multiple copies of a ligand in a single binding entity simultaneously forms multiple contacts with receptors, reducing the chance of complete dissociation of the ligand and increasing the probability of rebinding. Receptor clustering brings multiple binding ligands or receptors into close proximity allowing a multivalent process to take place.

The synergistic increase in binding affinity achieved by combining a number of monovalent ligands into a single multivalent species is classified as either cooperative or additive (Boas and Heegaard, 2004; Ercolani and Schiaffino, 2011). Additive binding is the result of increased binding efficiency achieved by having a greater number of binding entities per mole of ligand. In the additive case, the binding affinity is proportional to the monovalent binding affinity and the number of simultaneous independent binding interactions formed with a multivalent receptor. Cooperative binding can be either positively or negatively cooperative. In a positively cooperativity binding process, the observed affinity is stronger than would be expected for a purely additive process due to the action of linking monovalent ligands together into a single molecule. The converse of this is negative cooperativity, where the observed binding affinity is less than the expected additive interaction. Determining the degree of cooperativity is notoriously difficult and has led to a number of incorrect assessments of multivalent interactions (Ercolani, 2003). It should be noted that even in a negatively cooperative process, the binding affinity still exceeds the affinity of the corresponding monovalent interaction (Boas and Heegaard, 2004).

Biology nomenclature often describes receptors as the species present on the surface of a target cell or surface. In this investigation, the nomenclature recommended by Mammen *et al.* (1998) will be used, in which the receptor refers to ligand binding pockets on the protein and the ligand is any entity capable of binding to these pockets, regardless of whether the species is free in solution or immobilised on a solid phase.

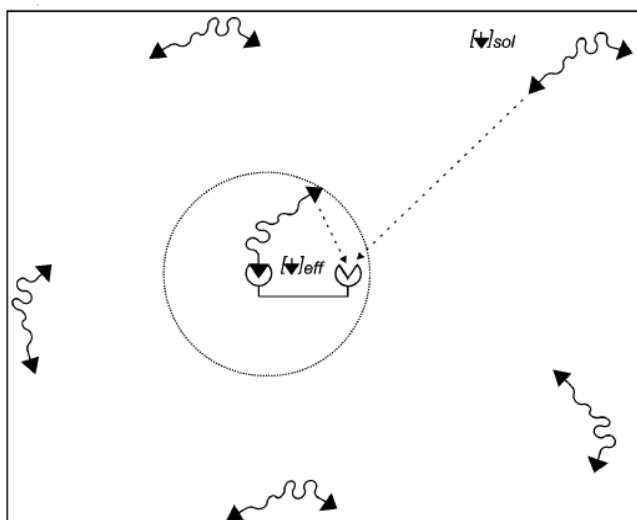
### 1.5.1. Thermodynamics of multivalent binding

The thermodynamics of a binding reaction determine the stability of the bound products relative to the unbound starting components. The relationship of factors affecting the stability of a protein ligand complex is given by the change in Gibbs free energy:

$$\Delta G = \Delta H - T\Delta S \quad \text{Equation 1.1.}$$

$\Delta G$  is the change in Gibbs free energy,  $\Delta H$  is the change in enthalpy,  $T$  is the temperature of the system and  $\Delta S$  is the change in entropy, also known as the change in the degree of disorder or the degrees of freedom. In thermodynamics, a process is spontaneous at a given temperature if it is accompanied by a negative change in the Gibbs free energy (Housecroft and Constable, 2006). The enthalpy of a multivalent binding reaction is proportional to the enthalpy of a single analogous monovalent binding event and the number of binding events occurring during the multivalent interaction. Entropy is a measure of the disorder within a system and is an aggregation of the translational, rotational and conformational entropy of an entity. A multivalent ligand is expected to have the same rotational and translational entropy as the comparable monovalent species, but greater conformational entropy due to the increased structural complexity of the linker used to combine the monovalent ligands into a single entity (Mulder *et al.*, 2004). This also leads to a greater reduction of conformational energy upon binding and has led some to suggest that minimising the degrees of freedom by introducing greater rigidity into the structure of the linker is important in the design of multivalent ligands (Mammen *et al.*, 1998a; Mammen *et al.*, 1998b; Krishnamurthy *et al.*, 2006). However, more recent evidence has found that flexible linkers retain a certain degree of flexibility upon binding, resulting in a smaller negative change in entropy and making conformational entropy less influential on the stability of multivalent complexes than expected (Krishnamurthy *et al.*, 2007). This is in agreement with the strong binding affinities observed for a number of multivalent ligands with flexible linkers (Kramer and Karpen, 1998; Mammen *et al.*, 1998; Kitov *et al.*, 2000; Choi, 2004). It has been argued subsequently that rigid structures can prevent ligands adopting the ideal conformation for binding, reducing the interaction's affinity (Deniaud *et al.*, 2011)

The conformational entropy is determined by the size of the multivalent scaffold and the length and nature of the linkers of which it is composed. The size and nature of the multivalent scaffold will often be determined by the physical properties of the targeted multivalent receptor, in particular the distance between the binding sites. The effect of the linker length and flexibility is explained by the concept of effective concentration. Effective concentration is the probability of two inter-linked entities reacting or forming a binding interaction (Kuhn, 1934; Jacobsen and Stockmayer, 1950). Mulder *et al.* (2004) give a clear example of effective concentration using a bivalent ligand in which one headgroup is already bound to the macromolecule (Figure 1.10).



**Figure 1.10.** Schematic of effective concentration of a partially bound multivalent ligand as depicted by Mulder *et al.*, (2004).  $[\blacktriangledown]_{\text{eff}}$  is the effective concentration of the ligand experienced by the unoccupied binding site of the partially bound receptor molecule.  $[\blacktriangledown]_{\text{sol}}$  is the real concentration of the ligand in the solution.

After the first headgroup has bound, the remaining unbound headgroup is essentially tethered close to the surface of the macromolecule. If the tether is of sufficient length and flexibility the second headgroup will be able to bind to the macromolecule without dissociation of the first. This gives a local concentration of the ligand that is higher than the real concentration, when no ligand is bound. In relation to conformational entropy, the optimal length of the linker is one that spans the distance between the two binding sites. Too short and the effective concentration will not be increased relative to the real concentration and a second ligand molecule will be needed to fully saturate the binding sites of the macromolecule. If only just long enough, binding of the second headgroup



will induce steric strain reducing conformational freedom and favouring dissociation of the headgroup. Conversely if too long, the headgroup will occupy a larger radius around the first headgroup, lowering the effective concentration and increasing the conformational entropy of the molecule.

### 1.5.2. Kinetics of multivalent binding

The effective concentration is a concept that falls under the broader subject of reaction kinetics. Kinetics describes the rate and extent to which a particular product will be produced relative to its reactants. In multivalent binding this can be thought of as the extent to which a particular multivalent complex will form. The extent of multivalent complex formation of a single ligand interacting with a single protein can be expressed as:

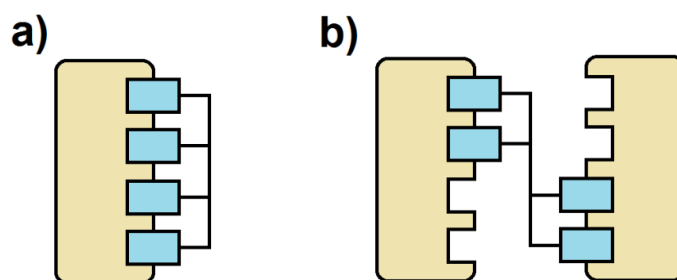
$$K_a = \frac{[ML]}{[M][L]} \quad \text{Equation 1.2.}$$

Where multivalent ligand, L, of concentration [L] interacts with macromolecule, M, present in concentration [M] to form complex ML present in concentration [ML] at equilibrium.  $K_a$  is the equilibrium constant of association for the complex. In a multivalent interaction, the equilibrium constant is the product of all the individual constants for each binding event between a single headgroup from the ligand and a single binding site of the macromolecule. The rate of the first interaction is expected to be slower than the rate for the monovalent process due to the greater complexity of the multivalent species (Chittasupho, 2012). But the binding of subsequent headgroups is expected to be faster due to the greater effective concentration and the increased probability of any one individual headgroup rebinding due to being tethered close to macromolecule binding site by the remaining bound headgroups of the ligand (Krishnamurthy *et al.*, 2007). The influence of the effective concentration on the overall binding constant is expressed by the relationship:

$$K_n = bK_1^n C_{eff}^{n-1} \quad \text{Equation 1.3.}$$

Where the overall association constant for a multivalent process involving  $n$  interactions is the product of a number of statistical factors related to the length and flexibility of the ligand,  $b$ , the association constant of a single headgroup attachment  $K_1$  and the effective concentration of a partially bound ligand,  $C_{\text{eff}}$ . The dissociation of the multivalent ligand is greatly reduced in comparison to the monovalent interaction because this requires dissociation at all points of attachment. Dissociation of headgroups is known to be stepwise and the probability of rebinding high (Rao *et al.*, 1998), so the concentration of species only attached to the macromolecule by a single point of attachment at any given time is a small fraction of total bound concentration. In contrast, the concentration of singly bound species in a monovalent interaction is equal to the total concentration of ML and once unbound the EC is equal to the real concentration of the ligand.

In addition to the determining the probability of dissociation, kinetics can be used to describe the probability of a multivalent event being intermolecular or intramolecular (Figure 1.11). When a multivalent ligand interacts with multiple binding sites from a single receptor molecule, the binding is classified as intramolecular, and when the single multivalent ligand interacts with multiple receptor molecules simultaneously the binding is considered to be intermolecular (Chittasupho, 2012; Figure 1.11).



**Figure 1.11.** a) Intramolecular and b) Intermolecular multivalent binding.

The preference of a multivalent ligand to form either an intra- or intermolecular species can be expressed by the ratio of the dissociation constant for the intermolecular reaction over the dissociation constant of the intramolecular reaction, which is commonly known as the effective molarity (EM; Equation 1.4).

$$M_{\text{eff}} = \left( \frac{K_n}{bK_1^n} \right)^{1/(n-1)} \quad \text{Equation 1.4.}$$

At low concentrations of ligand, intramolecular binding is favoured, but when the concentration of free ligand exceeds the effective concentration intermolecular interactions become favourable. While effective concentration is a theoretical value based on ligand geometry, size and flexibility, EM is quantified using the observed binding constants for the multivalent individual processes (Equation 1.4.; Mulder *et al.*, 2004). In this way, EM can measure the affinity enhancement over a monovalent process (Anderson *et al.* 1995; Rao *et al.*, 1998; Felluga *et al.*, 2000). EM has also been used to determine the ease of cyclisation in a multivalent system (Kirby, 1980; Galli and Mondolini, 2000), i.e. the ease of complete saturation of a single multivalent receptor by a single multivalent ligand. This has direct relevance to SAP, as Ho *et al.* (2005) have conducted investigations of complex formation under different concentrations of bivalent ligand and found the same ligand concentration dependence for the formation of decameric complexes of SAP using non-covalent, cross-linking, bivalent ligands. Furthermore, they found decameric complexes occur over a greater range of ligand concentrations for ligands with stronger binding affinities for SAP. Therefore, the design of decavalent ligands for SAP needs to consider the concentration of ligand when attempting to analyse protein-ligand complexes by methods such as x-ray crystallography, as the concentration of ligand will determine whether pentameric or decameric complexes are observed.

It is clear that the design of multivalent ligands is shaped by two factors, firstly the mode of binding sought, either intramolecular or intermolecular. Secondly, the effort to produce the most effective multivalent binding species for target receptor. The second factor may not always seek the ligand with the strongest binding affinity, greater specificity may also be important. Efficient multivalent binding requires a consideration of the individual binding process, the relative positioning of the binding sites and nature of the linker used to join multiple monovalent headgroups. The concepts of effective concentration, effective molarity and entropy are important in quantifying the influence of the multivalent ligand structure; while separate investigations of the monovalent binding interaction are required to determine the expected affinity gain and inform aspects of multivalent ligand design, such as the correct chemical linkage needed to ligate the monovalent headgroups into a single binding entity.

### 1.5.3. Examples of multivalent processes

There are many examples of multivalent binding processes in biological systems, but often they are not fully understood due to the complexity of the process, difficulty in identifying the exact constituents involved and cited examples of misinterpretation of binding cooperativity (Ecrolani, 2003).

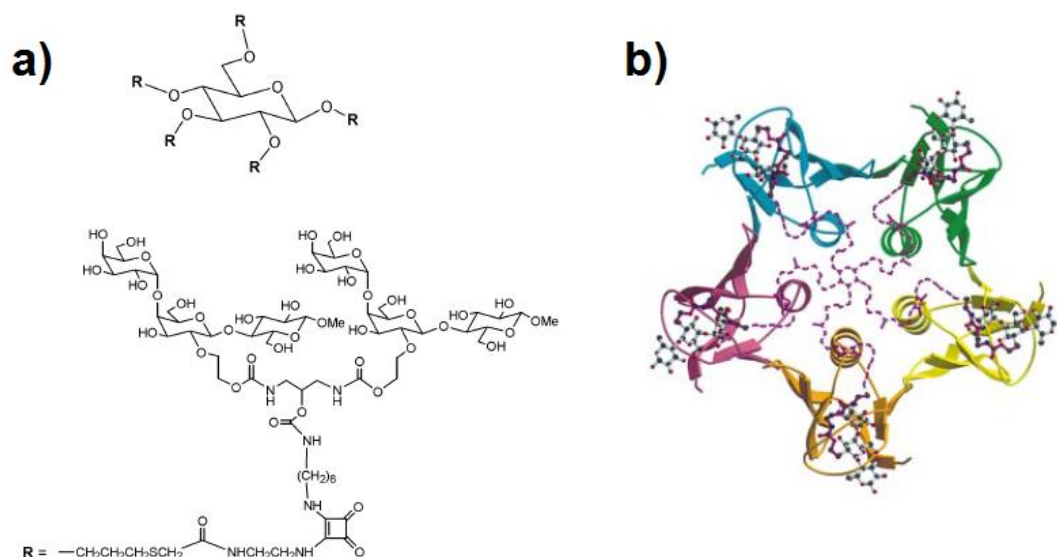
Concanavalin A (ConA) is a carbohydrate binding lectin that is probably one of the best characterised multivalent receptors (Goldstein *et al.*, 1965; Edelman *et al.*, 1972). For this reason it has been used as a model for numerous studies of multivalent ligand design and application (Kanai *et al.*, 1997; Dam *et al.*, 2000; Cairo *et al.*, 2002; Deniaud *et al.*, 2011). Similarly, its avid binding to carbohydrate ligands means ConA is widely used in the purification of glycoproteins (Iscoe *et al.*, 1974). Significant similarities have been observed between the tertiary structures of SAP and ConA subunits (Figure 1.12; Emsley *et al.*, 1994). In particular, both proteins have a flattened  $\beta$ -jellyroll topology and double cation binding sites. ConA induces agglutination of various cell types by interaction with carbohydrate moieties exposed on the surface of cells (Podolsky *et al.*, 1974). This interaction can be taken as an example of an intermolecular multivalent binding process in which ligands from multiple cells are bound by a single receptor, resulting in agglutination.



**Figure 1.12.** Cartoon representation of a subunit of Concanavalin A (PDB file: 3CNA). There are similarities in the arrangement of the  $\beta$ -sheet (yellow) and the double cation binding site. Loop regions are denoted by the green ribbon; calcium ion (limon yellow), manganese ion (violet).

A possible, example of multivalent binding relevant to the pentraxins is the proposed mechanism of complement activation by CRP during an acute phase response (Kaplan and Volanakis, 1974). CRP recognises phosphatidylcholine moieties presented on the surface of damaged cells and phosphocholine presented on the surface of some bacterial cells (Volanakis and Kaplan, 1971; Volanakis and Wirtz, 1979). The process is likely to be multivalent, with the strength of binding of CRP linked to the density of ligands on the entities surface. At a certain density, CRP will bind to the surface in sufficient density and distribution for the C1q complex of the complement system to bind, activating the complement system by the classical pathway (Peisajovich *et al.*, 2008). C1q consists of multiple globular headgroups which bind individually to CRP pentamers through a cleft on the A-face of the pentraxin (Gaboriaud *et al.*, 2003). Therefore, the proposed role of CRP in mediating the removal of nuclear material during cell death could be the result of two sequential multivalent binding processes. It could be contemplated that the need for two multivalent processes for the activation of complement reduces the possibility of improper complement activation. Although the exact groups involved and the mode of binding between CRP and C1q have yet to be elucidated, this example shows how the Pentraxins characteristic pentavalent structure could facilitate multivalent binding interactions as part of a function in the innate immune system.

An oft cited example of multivalent binding is the binding of Shiga toxins from *E. coli* to glycolipids on the surface of host cells. Shiga toxins are composed of six components in an AB<sub>5</sub> structure where the B<sub>5</sub> component has a pentameric symmetry (Figure 1.13; Ling *et al.*, 1998). Shiga-like toxins are responsible for millions of deaths from bacterial infections every year making it a popular drug target (Holmgren and Svennerholm, 1992). Monovalent therapeutics bind with only modest affinities, but multivalent analogues display greatly enhanced affinities capable of inhibiting binding to cell surfaces (Kitov *et al.*, 2000). Kitov *et al.* used a monosaccharide as the central core and alkyl chains to link five monovalent trisaccharide ligands into a single binding entity. The wide ranging potential applications means that a plethora of scaffolds have been developed for multivalent ligands.



**Figure 1.13.** a) The STARFISH Shiga-like toxin inhibitor molecule developed by Kitov *et al.*, 2000. b) STARFISH bound to the B5 subunit of Shiga-toxin 1 (SLT1) in an arrangement suggested by the authors. Cartoon representation of SLT1 with each subunit denoted by different colour. Figure adapted from Kitov *et al.*, 2000.

#### 1.5.4. Design of multivalent ligands

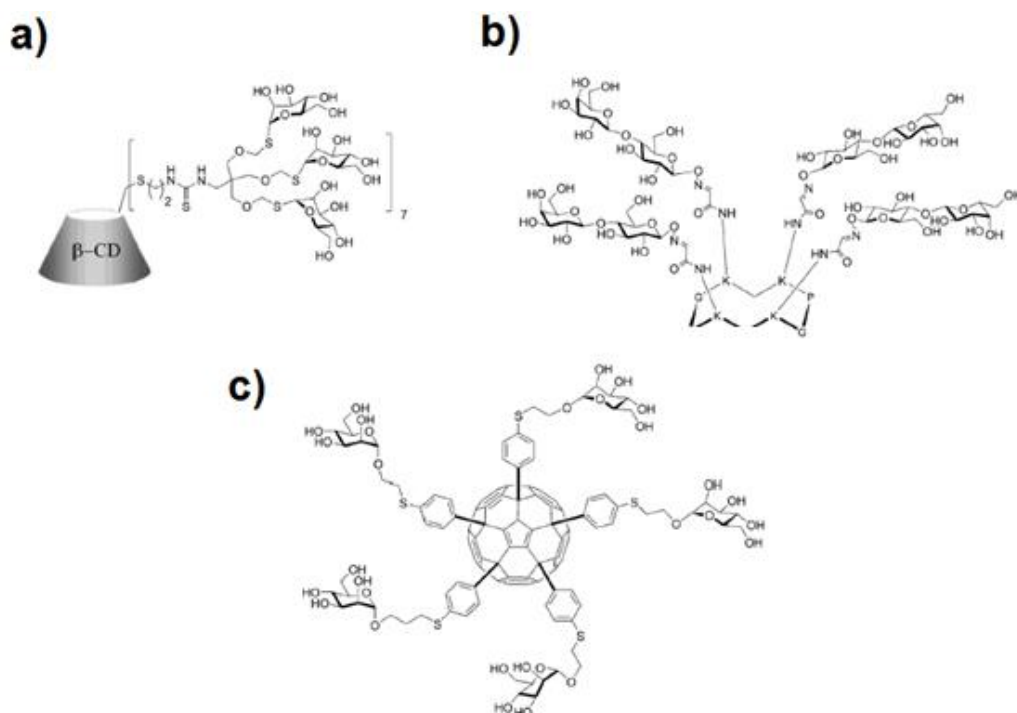
Multivalent ligands need to present multiple headgroups in an environment and orientation that is amenable to simultaneous binding by receptors on the surface of the target protein. Typically, this requires a scaffold to which the headgroups can be chemically-linked without modification of the key functional groups of the monovalent analogue. However, there are examples of self-assembling multivalent scaffolds, which maintain their structure by non-covalent interactions between building blocks such as peptides, DNA and micelles (Matsuura, 2013). There are also examples that combine the two methodologies, such as the use of liposomes to present multiple copies of multivalent saccharide ligands developed by Qu *et al.* (2014).

As discussed in the previous section, the design of a multivalent ligand should allow adequate size and flexibility to bridge the distance between multiple binding sites, so as to minimise steric strain. The design will also be significantly influenced by whether the ligand is intended to bind through an intra- or intermolecular mechanism. For intramolecular binding interactions targeting a 1:1 binding ratio, multivalent ligands should be designed to maximise the effective concentration, to discourage intermolecular binding and increase the probability of rebinding.

As a result of the design considerations above, multivalent ligands typically consist of a central core with linkers of specified length emanating outwards to a headgroup at their extremities capable of interacting with a binding site from the target. The number of headgroups is determined by the number of binding sites targeted and the length of the linkers dependent on the distance between the binding sites. Flexibility is a consideration in the design of a multivalent ligand (Kane, 2010), and the vast majority of multivalent scaffolds found in the literature have flexible linkers, although the central core may be rigid (Mammen *et al.*, 1998; Boas and Heegaard, 2004).

Covalently-bound chemical architectures are most relevant for SAP, as bivalent forms of monovalent ligands linked with hexanoyl chains have already shown significant promise as therapeutics. Much of the focus in designing multivalent ligands is on choice of the central scaffold (Mammen *et al.*, 1998). It has been reported that rigidification of the central scaffold can be detrimental to the ligands function, as it prevents ligands adopting the ideal geometry for simultaneous interaction with all binding sites (Deniaud *et al.*, 2011). However, rigidity is known to be an advantage if it is used to fix the scaffold in the optimal geometry for binding (Vrasidas *et al.*, 2003). For this reason, a number of multivalent ligands use rigid central scaffolds or structures known to maintain a particular shape. As previously mentioned Kitov *et al.* (2000) developed Shiga-like toxin inhibitors, which have a central saccharide unit to maintain a roughly pentameric symmetry. Oligosaccharides, such as cyclodextrins also offer rigid structures with controllable substitution (Figure 1.14; Gomez-Garcia *et al.*, 2005).

Rigid architectures are not limited to saccharide-based structures. Fullerene-based structures are one interesting application in which the rigid covalent structure can be used to fix the arrangement of ligands displayed on the surface (Isobe *et al.*, 2003).



**Figure 1.14.** Examples of rigid central scaffolds utilised for multivalent ligands a) cyclodextrin based ligands developed by Gomez-Garcia *et al.* (2005), b) cyclic decapeptide with rigidity provided by proline residues, which induce fixed  $\beta$ -turns, Renaudet and Dumy (2003), c) multivalent ligand using a fullerene to maintain the desired pentameric structure developed by Isobe *et al.* (2003).

Peptides are another class of biomolecule used in rigid scaffolds for multivalent ligands, despite the fact that peptides are not typically thought of as having well-defined secondary structures. Renaudet and Dumy (2003) showed that careful placement of proline residues in the amino acid sequence of a cyclic peptide can produce a rigid architecture. In their investigation they synthesised a peptide with rigidity conferred by the presence of two  $\beta$ -turns to fix the conformation and ensure the four ligands were presented on the same face of the central scaffold. In contrast, Zhang *et al.* (2004) showed that it is not always necessary to specifically include residues known to increase rigidity. Zhang *et al.* synthesised multivalent Cholera toxin inhibitors synthesised cyclic peptides with repeating  $(AA_n-K)_5$  sequences that demonstrated a 100,000-fold affinity gain over the corresponding monovalent ligand. They also performed computer simulations to show that the molecules maintained an expanded cyclic structure. Furthermore, they concluded that it was not important to present ligands on the same face of the central scaffold providing the arms used to link the ligands to the central scaffold were sufficiently long and flexible.



The multivalent examples given so far target a specific number of binding sites. However, there are a number of structures, which aim to present the maximum number of ligand headgroups to increase the chances of a successful interaction and maximise the probability of statistical rebinding. Perhaps one of the most well-known examples are the multiple antigen presenting (MAP) peptides reported by Tam (1988) for use in synthetic peptide vaccines. MAP peptides are more an example of dendrimers, which are a popular basis for presenting a high density of ligands (reviewed by Boas and Heegaard, 2004).

The chemical nature of the linker does not generally garner as much attention as the scaffold, instead the length is considered the critical feature (Deniaud *et al.*, 2011). Polymeric materials such as polyethylene glycol are common constituents of linkers, most likely due to their increased solubility over simple alkyl chains (Deniaud *et al.*, 2011); although, there are also many examples using diamino-based polymeric materials, particularly in the synthesis of dendrimer multivalent ligands (Boas and Heegaard, 2004). The Starburst<sup>TM</sup> dendrimers developed by Tomalia *et al.* (1985) are a classic example and also show the combination of diamino compounds and amino acids, specifically  $\beta$ -alanine. Therefore, research of the literature supports a flexible linker, with suitable solubility properties, which does not engage in specific interactions with the target and can be ligated to both the headgroup and a central scaffold. The headgroup should provide binding specificity and a moderate affinity, both of which will be enhanced in a multivalent ligand. Finally, the central scaffold should be relatively rigid that maintains the ideal geometry for multiple simultaneous interactions with the target's binding sites.

#### **1.5.5. Multivalency for SAP**

The structure of the Pentraxins look almost as if they are the result of natural receptor clustering with multiple identical receptors presented in a pentameric symmetry with strong non-covalent interactions between protomers (Emsley *et al.*, 1994). The current theory of SAP binding to amyloid deposits also relies on clustering of multiple monovalent ligands presented in configuration amenable to multivalent recognition by SAP (Pepys *et al.*, 1997). SAP does not bind to amyloid proteins in their native, soluble state. This may be due to the fact that the refolding needed for amyloid proteins to form

their characteristic cross- $\beta$  structure exposes a new motif that SAP recognises. Alternatively, it may simply be that for SAP recognition, a motif conserved after refolding of the amyloid protein needs to cluster with other SAP binding motifs to allow binding to occur with sufficient affinity.

Considering the homopentameric structure of SAP, it is difficult to contemplate binding to amyloid deposits occurring in anything other than a multivalent manner. However, this does not mean that SAP is binding the same ligand in each binding site or even that all binding sites are occupied. SAP is known to bind to a range of molecules found in amyloid deposits and the exact composition of the binding motifs for SAP may vary within the same amyloid deposit, let alone between deposits of different amyloid proteins (Pepys *et al.*, 1997).

## **1.6. Project Aims**

The aim of this investigation was to synthesise a range of multivalent ligands for SAP that exploit the expected cooperative binding of the protein's five identical subunits. The novel non-covalent cross-linking of SAP by CPHPC has been shown to have a greatly increased binding affinity over monovalent N-acetyl D-proline, so linking five CPHPC into a single molecule capable of simultaneous interaction with ten subunits from two SAP pentamers was also of significant interest as this could further improve the binding affinity. Therefore, the ligands would need to present pentameric symmetry, but with a flat structure that does not prevent close association of two pentamers with a multivalent ligand in the middle. This thesis documents the synthesis and characterisation of multivalent ligands for SAP conducted between 2010 to 2014. It also describes x-ray crystallographic studies of monovalent ligand N-acetyl D-proline binding to SAP to understand the exact interaction of the chosen headgroup for the multivalent ligands. Additionally, the x-ray crystal structure of Rat C-Reactive Protein (rCRP) is presented in complex with bivalent ligand CPHPC. rCRP has been the focus of a number of previous crystallographic studies due to its unusual structural features compared to other pentraxins. However, it has not been possible to obtain satisfactory data until now, and it is theorised that the use of novel bivalent ligands was a significant factor in overcoming previous difficulties.

## **Chapter 2. X-ray crystal structure of Serum Amyloid P component in complex with N-acetyl D-proline**

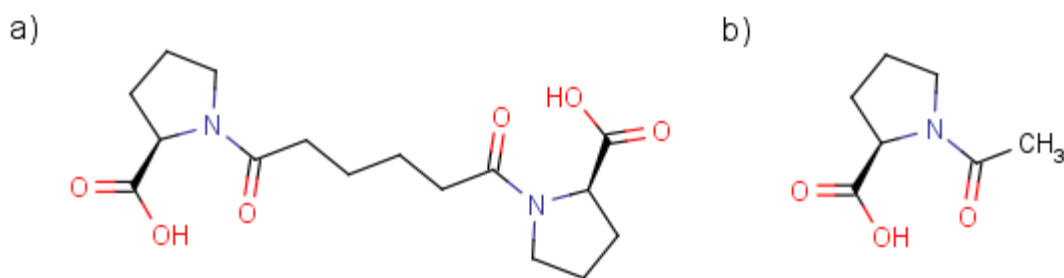
### **2.1. Introduction**

Over the past two decades significant efforts have been made to discover small molecule ligands capable of binding to SAP in a calcium dependent manner. Since the protein's discovery, it has been known that SAP is prevalent in amyloid deposits (Cathcart *et al.*, 1967). In 1995, Tennant and co-workers identified SAP's role in stabilising amyloid fibres and preventing proteolysis. SAP coats amyloid fibres, preventing recognition and removal by the immune system. This makes therapeutics capable of stripping SAP from amyloid deposits desirable drug targets in the treatment of amyloidosis.

No amyloid motif capable of binding to SAP has yet been identified, but numerous small molecule ligands for SAP have been identified and there are several crystal structures reported in the literature of SAP co-crystallised with a range of ligands including acetate, phosphoethanolamine, MO $\beta$ DG (Emsley *et al.*, 1994), dAMP (Hohenester *et al.*, 1997), N-acetyl D-proline (Purvis thesis, 2002), N-acetyl L-proline (Kolstoe *et al.*, 2014), CPHPC (Purvis thesis, 2002; Jenvey thesis, 2006; Kolstoe *et al.*, 2014).

CPHPC is a bivalent ligand capable of non-covalently cross-linking two SAP pentamers to form a decamer (Pepys *et al.*, 2002). CPHPC is currently under development as a treatment for amyloidosis (Kolstoe and Wood, 2010). Sustained treatment of patients with CPHPC results in sustained >95% depletion of circulating SAP by hepatocytes (Gillmore *et al.*, 2010). Despite binding to SAP with a low nanomolar dissociation constant (Kolstoe thesis, 2005), CPHPC cannot remove SAP bound to amyloid deposits. So an anti-SAP antibody can be used to locate directly to the amyloid deposit and highlight its presence to the immune system, which rapidly clears the mis-folded protein (Bodin *et al.*, 2010). Understanding the mechanics of the binding interaction between CPHPC and SAP is essential for intelligent development of treatments for amyloidosis, and for the design of multivalent ligands that mimic the action of CPHPC. Crystal structures of SAP with CPHPC have been attained (Purvis thesis, 2002; Jenvey thesis,

2006), but there is some ambiguity surrounding the nature of the ligand interaction with amino acid residues adjacent to the calcium binding site.



**Figure 2.1.** a) Bis D-proline compound, CPHPC. b) N-acetyl D-proline, NADPro.

SAP binds to small molecule ligands in a calcium-dependent manner. This interaction is further improved when the small molecule has a non-polar functional group capable of fitting into the small hydrophobic pocket adjacent to the calcium binding site. The D-proline headgroups of CPHPC bind with the pyrrolidine ring fitting tightly into the hydrophobic pocket formed by Leu62, Tyr64 and Tyr74 from SAP (Purvis thesis, 2002). However, it is unclear whether the binding interaction extends beyond the calcium binding site and the hydrophobic pocket. The carbonyl group of the hexanoyl chain in CPHPC is known to be in close proximity to the phenyl side chain of Tyr64, but there is no conclusive evidence in the literature that demonstrates the presence or absence of a hydrogen bond between the two functional groups.

N-acetyl D-proline (NADPro) is not restrained by the hexanoyl chain that is present in CPHPC (Figure 2.1), therefore the crystal structure of NADPro in complex with SAP should show if any preference exists between the *cis* and *trans* conformation for the carbonyl oxygen of the acetyl group when the ligand is unrestrained by other functional groups. A low resolution crystal structure of NADPro has previously been collected but as it is limited to a resolution of just 2.4Å (Purvis thesis, 2002), the electron density map is not sufficiently detailed to discern if any preference exists between the *cis* and *trans* isomers of the ligand's acetyl group. For this purpose, a high resolution structure of NADPro in complex with SAP was sought. If the acetyl is shown to form a hydrogen bond with SAP then it would be beneficial to formulate CPHPC in the *cis/cis* conformation such that any energy barrier to the binding of SAP is minimised, alternatively it may be beneficial to design a new compound to exploit this interaction. Furthermore, a detailed understanding of the nature of SAP ligand interaction is crucial

Chapter 2. X-ray crystal structure of Serum Amyloid P component in complex with N-acetyl D-proline

to design multivalent ligands capable of binding to SAP with the maximum binding affinity.

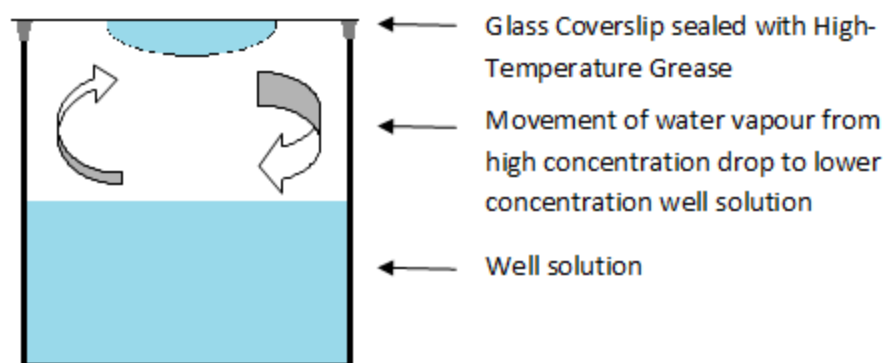
## **2.2. Methods and Results**

### **2.2.1. Crystallisation of SAP with N-acetyl D-proline**

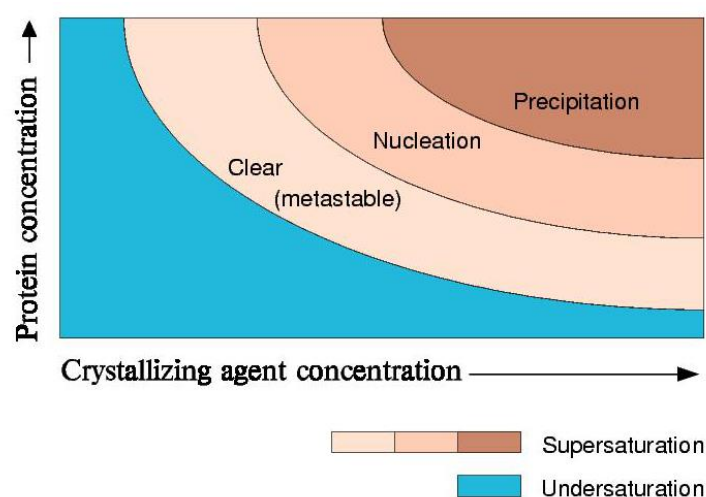
SAP was purified from human sera in Professor Sir Mark Pepys's laboratory (London, UK) according to the procedure outlined by Hawkins and co-workers (1991). In brief, SAP is isolated from sera of healthy donors by calcium-dependent affinity chromatography using a phosphoethanolamine (PE) sepharose column. The column is equilibrated with TC buffer (10mM Tris, pH 8.0, 140mM NaCl, 2mM CaCl<sub>2</sub>, 0.1% w/v NaN<sub>3</sub>), followed by elution with TE buffer (10mM Tris, pH 8.0, 140mM NaCl, 10mM EDTA, 0.1% w/v NaN<sub>3</sub>). SAP is then concentrated by ultrafiltration and dialysed against TN buffer (10mM Tris, pH 8.0, 140mM NaCl, 0.1% w/v NaN<sub>3</sub>).

N-acetyl D-proline (NADPro) was purchased from Bachem AG (Bubendorf, Switzerland). SAP was co-crystallised with NADPro by hanging drop vapour diffusion at 4°C. In this method, microlitre drops of protein-ligand solution are mixed with similar volumes of a crystallisation cocktail on a siliconised coverslip. The mixed drop of protein/cocktail solution is suspended by surface tension above a 1ml well of the crystallisation cocktail in a sealed environment (Figure 2.2). Crystallisation cocktails are designed to induce slow precipitation of a protein over time. Typically they contain a precipitant, such as polyethylene glycol, buffer, salts, additives and cryoprotectants. The protein drop contains a lower concentration of precipitant than the well solution, so water diffuses through the vapour phase from the drop to the well solution. As the drop shrinks, the concentration of protein slowly increases until it reaches a supersaturation point and precipitates from solution. If the rate of precipitation is gradual enough, then the protein can precipitate in a crystalline form; the protein solubility diagram in Figure 2.3 illustrates how the protein solubility changes with changing protein and precipitant concentration, as a vapour diffusion experiment proceeds, the volume of the solvent containing the protein shrinks. As the solvent volume decreases the protein concentration increases, ideally to the point that it reaches the nucleation phase of Figure 2.3, at this point crystallisation may begin, reducing the concentration of protein

Chapter 2. X-ray crystal structure of Serum Amyloid P component in complex with N-acetyl D-proline in the aqueous phase to the metastable point where crystal growth may continue.



**Figure 2.2.** Diagram of hanging drop vapour diffusion crystallisation experiment.



**Figure 2.3.** Protein solubility diagram. Graph demonstrates how the protein solubility changes with increasing concentration of protein and precipitant. During a vapour diffusion experiment, the volume of the aqueous solvent decreases in the protein drop, leading to an increase in both protein and precipitant concentration, ideally this brings the concentration of protein into the nucleation phase, at which point it begins to crystallise (Image taken from [http://www-structmed.cimr.cam.ac.uk/Course/Crystals/Theory/phase\\_diag\\_zones.html](http://www-structmed.cimr.cam.ac.uk/Course/Crystals/Theory/phase_diag_zones.html)).

Crystallisation of proteins is not trivial and it is not possible to predict the conditions required to produce crystals for x-ray diffraction experiments when dealing with a never-before crystallised protein. Typically, an effort to crystallise a protein *ab initio* utilises a sparse matrix screen in which 96 conditions used to successfully grow crystals of unrelated proteins are tested with the target protein (Page *et al.*, 2003). Methods for crystallising SAP are widely available in the literature, so conditions described by Purvis (thesis, 2002) were chosen to co-crystallise SAP with NADPro using a grid

screen to refine the published parameters. A grid screen typically alters two variables, such as precipitant concentration and pH, in a linear fashion centred around previously determined conditions (McPherson, 1982). Small differences in the crystallisation cocktail can greatly influence the quality of crystals produced. A grid screen is a useful way of varying crystallisation conditions in a narrow range to optimise “hits” from a sparse matrix screen. Crystals of NADPro-SAP were obtained using 10.1mg ml<sup>-1</sup> protein with a 20-fold molar excess of NADPro at 4°C with 15% w/v PEG 550 MME, 60mM Tris buffer, pH 7.5, 84mM NaCl, 10mM CaCl<sub>2</sub> and 18% v/v glycerol.

X-ray crystallography experiments are typically performed at 100K. Modern synchrotrons produce powerful radiation capable of generating radical species within the crystal, which can damage the protein, leading to rapid deterioration of the crystal at room temperature (Garman, 1999). Cooling crystals in a stream of cold nitrogen gas during data collection reduces the thermal motion of radical species produced by x-ray radiation. However, the cryo-cooling process can induce ice formation that destroys the protein crystal. It is often necessary to add a cryoprotectant to the crystallisation cocktail to prevent ice crystals forming. The cryoprotectant can be added directly to the crystallisation cocktail prior to the vapour diffusion experiment, or after crystallisation by soaking the protein crystal in a drop of well solution containing the desired cryoprotectant. In this experiment, crystals of SAP were obtained using a crystallisation cocktail that already contained 18% v/v glycerol and this proved sufficient to prevent ice formation upon cryo-cooling.

Crystals of SAP-NADPro were removed from the crystallisation drop in a small amount of mother liquor using a nylon loop. The loop was immediately plunged into liquid nitrogen and stored at 100K prior to data collection.

### **2.2.2. X-ray diffraction data collection**

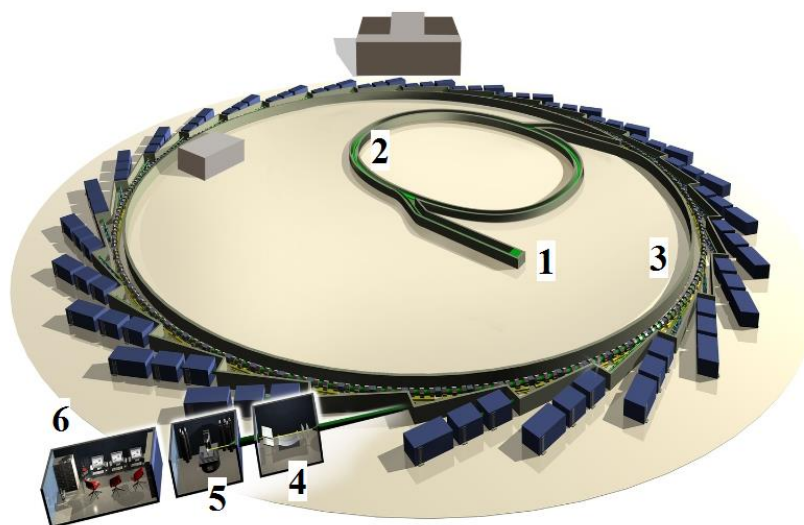
X-rays are electromagnetic radiation with wavelengths between 100Å to 0.1Å first identified by Wilhelm Conrad Röntgen in 1895. Henry and Lawrence Bragg were the first to realise that x-rays were scattered by the periodic arrangement of atoms in a crystal (Rupp, 2010). Furthermore, they discovered that the intensity and angle at which x-rays are scattered by a crystal is directly related to the position and identity of the

atoms within the crystal. By recording the diffraction pattern produced by a crystal from multiple different angles, it is possible to calculate the exact position of the atoms, or more specifically, pinpoint areas of electron density within the crystal.

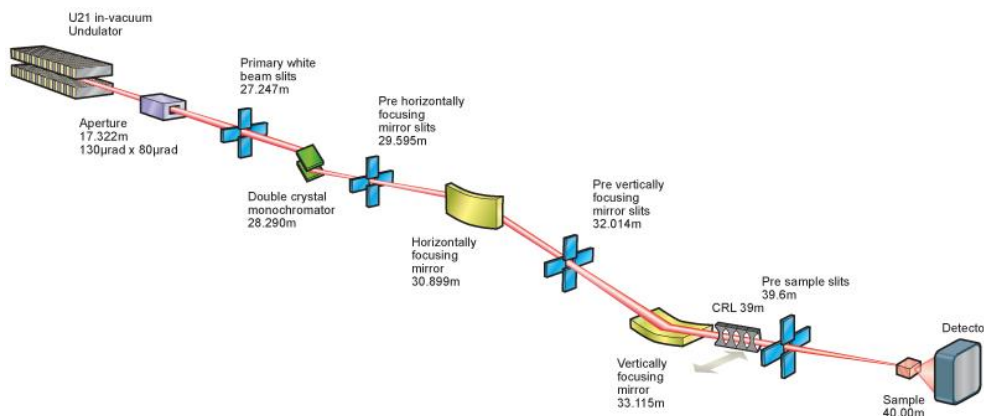
X-rays are absorbed by atoms, causing oscillations in electrons, which then emit radiation of the same wavelength as the absorbed radiation. Waves emitted from a single atom combine with those emitted from all other atoms in the unit cell. The unit cell of a protein crystal is the minimum repeating unit needed to replicate the entire crystal structure by repeated lattice translations. As the emitted radiation can be thought of as a 2-dimensional sine wave, the phase or relative displacement of the wave along the x-axis is important. Waves with the same displacement are “in-phase” and combine constructively, increasing the amplitude of the combined wave, whereas waves with opposite phases combine destructively, reducing the combined amplitude. Only diffracted x-rays where the partial waves combine in a constructive manner result in a reflection that appears as a spot on a diffraction pattern which can be recorded and measured by a detector. Placing a photon detector in-line with the x-ray source and the crystal allows accurate measurement of the intensity of diffracted waves that can be used in subsequent investigations to solve the crystal structure of a protein.

X-ray diffraction data was collected on beamline I03 at Diamond Light Source, DLS, (Didcot, UK). DLS is a third generation synchrotron which accelerates electrons in a vacuum to >99% the speed of light using electromagnets. Powerful electromagnets maintain the electron's speed and synchronous path (Figure 2.4; Rupp, 2010). As the path of the electron changes, it emits energy in the form of electromagnetic radiation. Insertion devices, such as wigglers and undulators, can be inserted into a synchrotron to produce an even more intense beam of radiation by causing rapid oscillation in the electron's path. The beam of radiation is passed through a series of mirrors and slits to produce a finely collimated beam of monochromatic x-rays, which are used by a beamline in diffraction experiments (Figure 2.5).





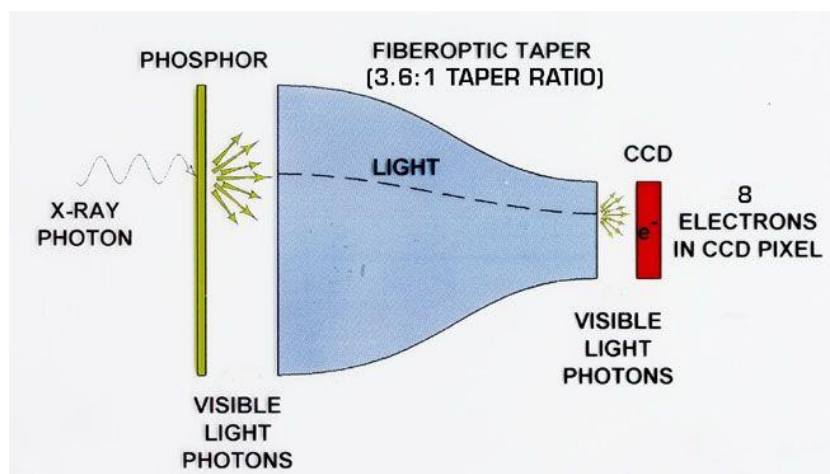
**Figure 2.4.** Diagram of Diamond Light Source Synchrotron (Didcot, UK). 1. Electron gun. 2. Booster ring. 3. Storage ring with bending magnets to maintain synchronous path. 4. Monochromator for selecting single wavelength of radiation. 5. Experimental hutch containing crystal mounting and data collection apparatus. 6. Beamline control room. Image adapted from: <http://www.diamond.ac.uk/PressOffice/MediaResources.html> on 27/04/2014



**Figure 2.5.** Diagram of a typical synchrotron beamline. Shows the undulator which stimulates emission of intense x-ray radiation by electrons as they pass through the insertion device. A series of mirrors and slits selects a single wavelength used in the diffraction experiment. Image taken from <http://www.diamond.ac.uk/Beamlines/Mx/I03/Beamline-Schematic.html> on 27/04/2014

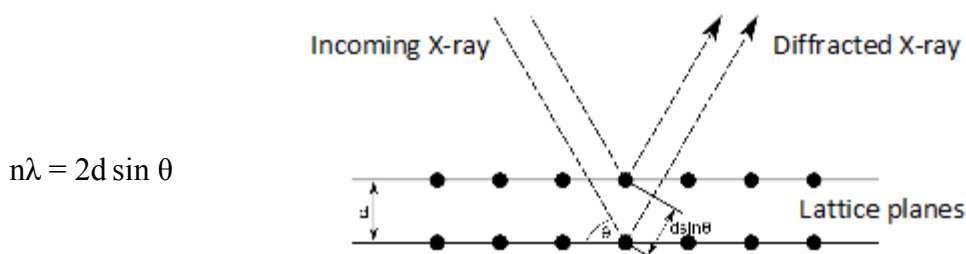
Upon exposure of the crystal to the x-ray beam, a diffraction pattern is produced, which can be recorded by a detector. In recent years, synchrotrons have replaced their charge-coupled device (CCD) detectors with Pilatus Hybrid Pixel detectors leading to a significant change in data collection methodologies (Brönnimann *et al.*, 2003). The x-

ray diffraction data for the SAP-NADPro crystals were recorded on a CCD detector, but subsequent chapters will present data collected using Pilatus detectors. CCD detectors consist of a thin fluorescent film that absorbs x-ray radiation and emits visible light, which is carried along a fiber glass taper leading to a CCD semiconductor array where free electrons are generated. This produces an electrical signal proportional to the number of photons detected, so the exact photon count can be stored digitally as a measure of the diffraction spot intensity.



**Figure 2.6.** Charge-coupled device x-ray detector. An x-ray strikes the fluorescent surface of the detector causing emission of visible light proportional to the number of electrons responsible for producing the detected x-ray, which travels down a fibre optic taper to a charge coupled device which measures the intensity of the visible light and stores the photon count for later analysis (Image taken from: [proteincrystallography.org](http://proteincrystallography.org) Date: 8/5/2014)

Monochromatic radiation of known wavelength is essential for crystallographic studies to facilitate the determination of atom spacing using Bragg's equation.



**Figure 2.7.** Left: Bragg's law where  $n$  is an integer,  $d$  is the spacing between diffraction scattering planes and  $\theta$  is the diffraction angle of the wave relative to the lattice plane. Right: representation of incoming x-ray being diffracted by two parallel planes of atoms within a crystal.

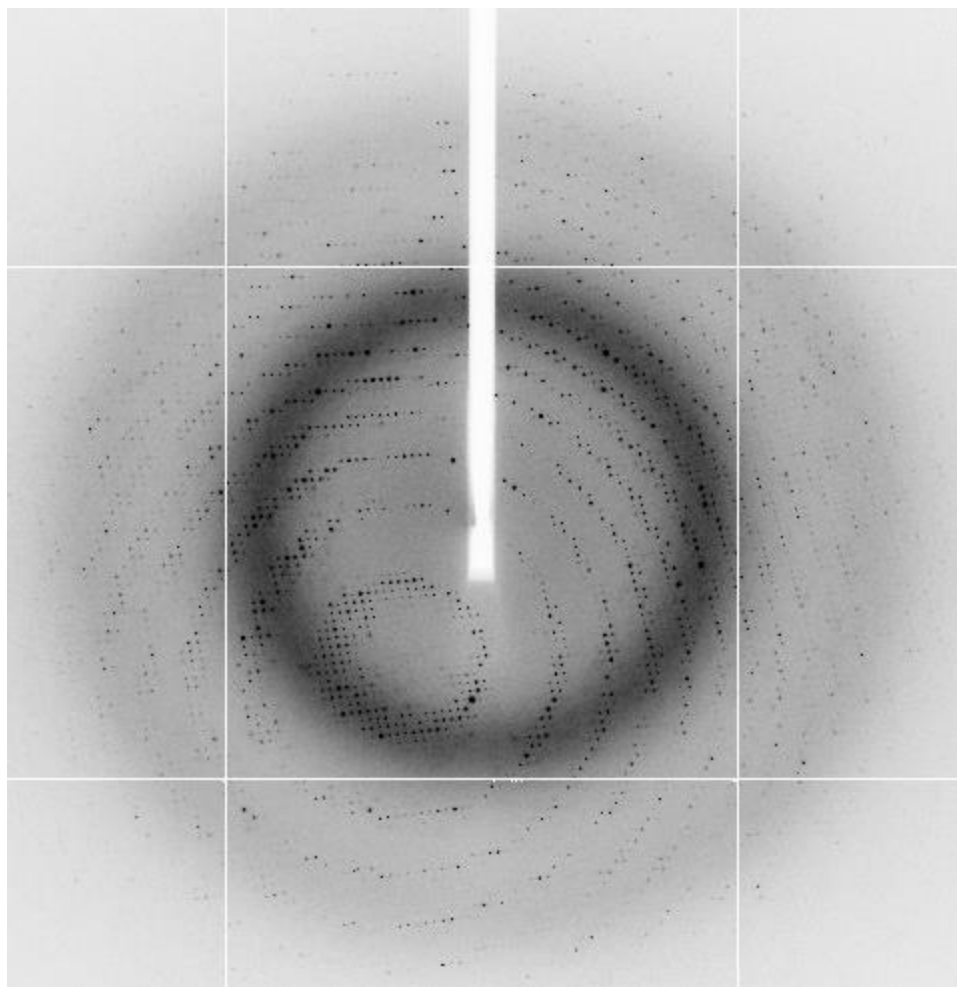
Bragg's law visualises reflections as occurring on sets of parallel lattice planes within a crystal. As previously discussed, diffracted waves emanating from the crystal must be in-phase for them to combine constructively and result in a signal at the detector for a particular reflection. Bragg's law defines the interplanar spacing necessary for two waves of wavelength,  $\lambda$ , and diffraction angle,  $\theta$ , to produce maximum constructive interference. Each spot recorded by the detector is the combined wave produced by the diffraction contribution of every atom in the unit cell. The structure factor,  $F_{hkl}$ , is a complex function that can be calculated by Fourier summation of the diffraction from all atoms in the unit cell with a magnitude proportional to the square root of the intensity of a diffraction spot (Rupp, 2010).

$$F_{hkl} = \sum_{j=1}^n f_j e^{2\pi i(hx_j + ky_j + lz_j)} \quad \text{Equation 2.1}$$

The structure factor equation sums the diffraction contributions of all atoms, where  $f_j$  is the scattering contribution of atom  $j$ .  $i$  is the imaginary number  $(-1)^{1/2}$ ,  $x$ ,  $y$ ,  $z$  are the fractional coordinates of the atom in the unit cell and  $h$ ,  $k$ ,  $l$  are the coordinates of a reflection in the reciprocal lattice produced by a set of parallel planes within the crystal. The atomic scattering factor,  $f_j$ , treats atoms as spherical regions of electron density where the magnitude of the contribution to the overall structure factor is related to the number of electrons it possesses. Therefore, the amplitude of the diffracted wave from one atom is determined by the number of electrons. The exponential part of the equation shows the fractional displacement of atom  $j$  along each unit cell axis, indicating that the phase is related to the position of the atom within the unit cell. The amplitude of the waves that produced a particular reflection on the diffraction pattern can be obtained from the data because it can be shown that the square root of the intensity recorded for a spot on the diffraction pattern is proportional to amplitude. However, the phase is not recorded by the diffraction experiment and therefore must be determined separately.

### 2.2.3. Data quality assessment

Once the crystal had been loaded onto the synchrotron beamline, two diffraction images separated by a  $90^\circ$  rotation of the crystal were collected to determine the spacegroup of the crystal and whether the crystal could produce diffraction data of sufficient quality for data collection. The diffraction pattern of the first x-ray exposure is pictured below.



**Figure 2.8.** Diffraction image of NADPro-SAP crystal NAD3.

Manual inspection of the diffraction pattern indicated that the crystal diffracted to a resolution of at least  $1.5\text{\AA}$  with the indexing program Mosflm determining the mosaicity to be  $0.77^\circ$  (Powell, 1999). The maximum resolution is the minimum interplanar spacing for which reflections can be recorded and mosaicity refers to the relative displacement of repeating units within the crystal. The crystal can be thought of as a 3-dimensional mosaic of unit cell blocks, which when well aligned produce sharp, narrow peaks on the diffraction pattern. If however the unit cells are not well aligned and the

mosaic spread is high then a single reflection may persist over a number of sequential diffraction images. The peak profiles in a crystal with high mosaicity appear smeared making it difficult to accurately determine the maximum peak height and index the reflection. The mosaicity of the NAD3 crystal was less than  $1.00^\circ$  so was considered usable for x-ray diffraction. Processing of the two images with Mosflm indicated a  $0.5^\circ$  oscillation angle was required for each frame to prevent reflections overlapping onto multiple frames. In total 360 images were collected using 25% transmission of the synchrotron beam and a  $0.5^\circ$  oscillation angle to record  $180^\circ$  of diffraction data with a maximum resolution of  $1.4\text{\AA}$ .

#### 2.2.4. Data processing

The raw diffraction data collected from DLS was processed using Mosflm. The first stage of data processing is to index reflections from the diffraction pattern and assign the unit cell of the crystal. A unit cell's dimensions are given by axes  $a$ ,  $b$  and  $c$  and angles  $\alpha$ ,  $\beta$ , and  $\gamma$ . Each unit cell can be assigned to 1 of 14 Bravais lattices based on their dimensions and translational symmetry (Rupp, 2010). Symmetry operations such as rotations and screw axes allow identical motifs within the unit cell to be superimposed upon each other. The asymmetric unit is the smallest motif in the unit cell that can be used to replicate the entire unit cell contents through its symmetry operations. Combining all possible symmetry operations with the 14 Bravais lattices gives rise to 65 spacegroups in which a protein can crystallise. Mosflm and other data processing programs are used to determine the spacegroup in which the protein crystallised.

Each set of planes within the unit cell that act as sources of diffraction, as defined by Bragg's law, produce a reflection at a particular angle. The concept of the reciprocal lattice can be used to determine the position at which a particular reflection will be detected. Each real-space lattice point is related to a corresponding point in the reciprocal lattice. If a vector from an arbitrary origin with coordinates  $(0, 0, 0)$  to the real-space lattice point has a magnitude of  $d_{hkl}$ , then magnitude of the corresponding reciprocal lattice point will be  $1/d_{hkl}$ . Reciprocal lattice points are produced by diffraction from a set of parallel planes within the real-space lattice. Miller indices,  $hkl$ , indicate the frequency with which a set of planes intersect a particular axis, for example

a set of planes with miller indices (5, 3, 4) intersects the a-axis 5 times, the b-axis 3 times and the c-axis 4 times. Each set of parallel planes produces one spot which can be assigned coordinates hkl in the reciprocal lattice. Mosflm can calculate the unit cell dimensions by treating reciprocal lattice points as vectors between the origin and each reflection (Powell, 1999). To determine the spacegroup, Mosflm selects approximately 200 reflections from 1 or more diffraction images using a peak search routine that includes choosing reflections with suitable signal-to-noise,  $I/\sigma(I)$ , ratio. The selected peaks are then used to compute the smallest triclinic cell with minimal symmetry. The basic cell is then compared to the 64 other possible space groups by distorting the unit cell dimensions and comparing the expected position of reflections for a given space group against those observed in the diffraction pattern. A penalty score is applied to each unit cell based on the results, with the space group of the highest symmetry, but lowest penalty generally selected. Table 2.1 shows the results of the spacegroup determination routine in Mosflm. The solution before the first big increase in penalty scores is generally selected as the correct cell symmetry, in this case the monoclinic primitive spacegroup from solution 3.

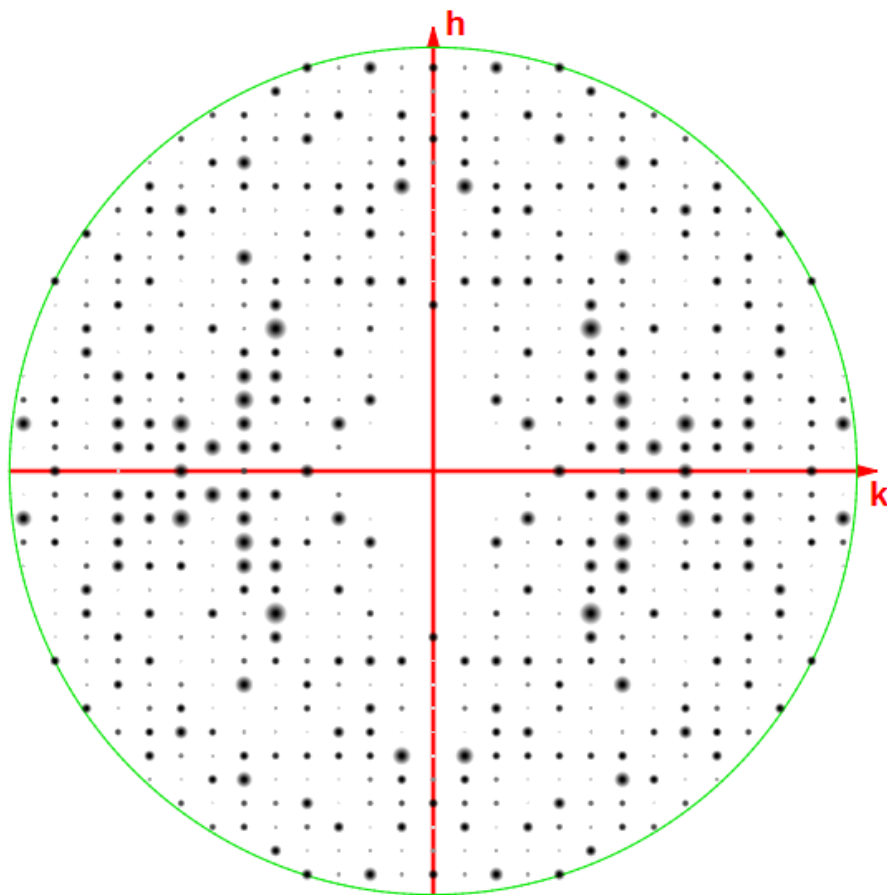
Solution	Lattice	Penalty	Unit cell dimensions					
			a (Å)	b (Å)	c (Å)	$\alpha$ (°)	$\beta$ (°)	$\gamma$ (°)
1	aP	0	71.1	94.9	99.7	96.9	89.9	90.1
2	aP	0	71.1	94.9	99.7	83.1	89.9	89.9
3	mP	1	94.9	71.0	99.7	90.0	96.9	90.0
4	mC	41	146.0	129.1	71.1	90.0	90.2	90.0
5	oC	42	129.0	145.6	71.2	90.0	90.0	90.0
6	mC	42	128.9	145.6	71.1	90.0	90.2	90.0

**Table 2.1.** Mosflm unit cell determination results for SAP-NADPro crystal prior to full data collection. Chosen solution highlighted in blue.

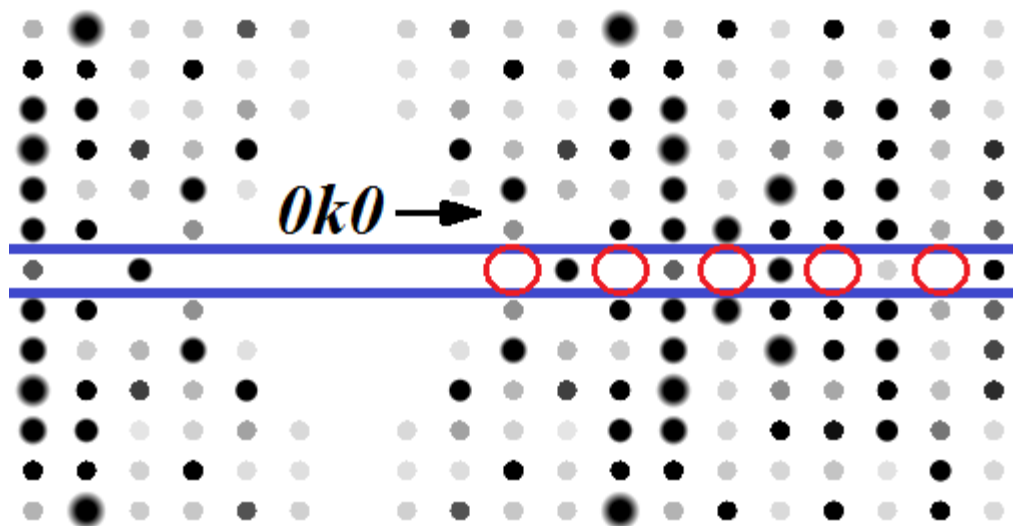
P2<sub>1</sub> was indicated as the correct spacegroup based on the results of the indexing procedure, with unit cell dimensions of  $a = 94.42\text{Å}$ ,  $b = 70.56\text{Å}$ ,  $c = 99.06\text{Å}$   $\alpha = \gamma = 90.00^\circ$   $\beta = 96.90^\circ$  after cell refinement. The data was integrated in Mosflm to estimate the intensity of each reflection with an associated error after subtracting the background noise. Integration is achieved by drawing a box around the area of a diffraction image

where a reflection is predicted by the spacegroup and measuring the intensity of the peak while subtracting the background signal recorded by pixels neighbouring the reflection spot. This process reduces the data from many individual image files into a single reflection, MTZ, file.

The program ViewHKL (Evans, 2006) reads MTZ files and can be used to view the recorded reflections along different planes in the reciprocal lattice. Symmetry elements in the unit cell cause some reflections to be absent from the diffraction pattern. In the case of spacegroup  $P2_1$  the two-fold screw axis dictates that certain reflections have symmetry mates with exactly opposite phases, which leads to perfect annihilation of the signal and the absence of a reflection in the diffraction pattern. Each spacegroup has specific conditions for systematically absent reflections, for spacegroup  $P2_1$ , each reflection along reciprocal axis,  $k$ , where  $k = 2n$  will be absent (Kopský and Litvin, 2010). ViewHKL can be used to see if systematic absences are observed on the expected planes and thereby validate the choice of spacegroup. Figure 2.9 shows the  $hk0$  plane of the NADPro-SAP data and Figure 2.10 is an enhanced image showing the  $0k0$  plane with the expected position of the absent reflections circled. These figures show absent reflections in the expected positions, therefore it is highly likely that  $P2_1$  is the correct spacegroup for the NADPro-SAP crystal.



**Figure 2.9.** Precession image of NADPro-SAP diffraction intensities of the  $hk0$  plane.



**Figure 2.10.** Enhanced precession image viewing the  $0k0$  plane. Red circles show the systematic absences at intervals of  $k = 2n$  caused by symmetry elements of the  $P2_1$ .



### 2.2.5. Merging and Scaling

The integrated data were scaled and merged in SCALA (Kabsch, 1988) from the CCP4 suite of programs (Collaborative Computer Project 4; Winn *et al.*, 2011). During data collection a number of factors can cause the measured intensities to vary and as such they are not on the same scale (Evans, 2006). Factors affecting the recorded intensity include variation of incident x-ray beam intensity and radiation damage. Scaling compares symmetry related reflections to calculate and apply appropriate correction factors to minimise the difference between their measured intensities. Once scaled, it is possible to identify the resolution limit and any bad regions in the data. Internal consistency in the data can be measured with R-factors such as R-merge, R-rim and R-pim. R-merge, also known as R-sym, is the most inferior of the R-factors as it does not adjust for multiplicity, which is multiple measurements of the same symmetry-related reflection. Having multiple observations of the same reflection should result in increased precision of the measurement, but R-merge increases with increasing multiplicity (Weiss *et al.*, 1997). R-rim (a.k.a. R-meas) and R-pim correct for increased multiplicity and are therefore more instructive (Diederichs and Karplus, 1997).

The merging R-factors are defined as:

$$R_{merge} = \frac{\sum_h \sum_{i=1}^N |I_{(h)i} - \bar{I}_h|}{\sum_h \sum_{i=1}^N I_{(h)i}} \quad \text{Equation 2.2.}$$

$$R_{rim} = \sum \left( \frac{N}{N-1} \right)^{1/2} \frac{\sum_h \sum_{i=1}^N |I_{(h)i} - \bar{I}_h|}{\sum_h \sum_{i=1}^N I_{(h)i}} \quad \text{Equation 2.3.}$$

$$R_{rim} = \sum \left( \frac{1}{N-1} \right)^{1/2} \frac{\sum_h \sum_{i=1}^N |I_{(h)i} - \bar{I}_h|}{\sum_h \sum_{i=1}^N I_{(h)i}} \quad \text{Equation 2.4.}$$

The merging R-factors take  $N$  number of observations of the intensity  $I$  of reflection  $h$  and calculate the difference between each observation and the average recorded intensity for the reflection. A high R-merge value can also be an indication of incorrect indexing, an R-merge covering all reflections in excess of 60% could be an indication

that the wrong spacegroup has been selected. In this scenario, the reflections on the diffraction pattern would deviate from their expected location causing a high variability in the observed intensities. For NADPro-SAP, the R-pim averaged over all reflections was 3.6% further supporting the assignment of  $P2_1$  as the correct spacegroup.

SCALA outputs merging statistics for each resolution shell of the data, Table 2.2 provides a summary of the statistics for SAP-NADPro crystal NAD3. There is little agreement within the crystallographic community about the correct way to determine the resolution limit of diffraction data. But, a signal-to-noise ratio,  $I/\sigma(I) > 2.0$  in the highest resolution shell is good indicator of the strength of recorded reflections. R-factors can be calculated for reflections within different resolution ranges to give an indication of the internal consistency for particular parts of the diffraction pattern. Weaker, high resolution reflections invariably have higher merging R-values due to the low signal-to-noise ratio and this can be used to judge a potential resolution cut-off to prevent noise being included in the dataset. The NAD3 dataset shows an R-pim of 38.5% in the highest resolution shell with an  $I/\sigma(I)$  of 2.6 indicating that diffraction may have been detectable beyond  $1.5\text{\AA}$ , however this was not apparent during data collection and therefore, no attempt was made to collect reflections of higher resolution. An analysis of the resolution of crystallographic data for structures deposited in the protein data bank reveals ([www.rcsb.org/pdb](http://www.rcsb.org/pdb)) that the resolution of the NAD3 data is in the 90<sup>th</sup> percentile of all deposited structures, so it is not too concerning that the resolution limit is conservative. Overall completeness was acceptable, >95%, and the multiplicity is high at 7.0 considering the relatively low symmetry of spacegroup  $P2_1$ .

Parameter	Crystal NAD3
Space Group	P2 <sub>1</sub>
Unit Cell Dimensions	a = 94.42, b = 70.56, c = 99.06 Å $\beta$ = 96.90°
Resolution Range	28.66 – 1.50 Å
No. of Reflections	1 392 596
No. of Unique Reflections	198 644
Multiplicity	7.0 (6.9)
Completeness (%)	96.5 (91.0)
R <sub>pim</sub>	0.036 (0.385)
(I)/ $\sigma$ (I)	14.1 (2.6)
Wilson B-factor	14.6
<b>Refinement</b>	
Molecular Replacement Model	1SAC
No. of Residues	1020
No. of Solvent Molecules	962 (Water), 5 (GOL)
No. of Ligand Molecules	5 (NAD), 10 (Ca), 5 (NAG)
Rmsd Bonds	0.007Å
Rmsd Angles	1.26°
Residues in Favoured Regions (%)	98.17
Residues in Disallowed Regions (%)	0
<b>Average B-factor</b>	
Protein	14.72
Ligand	24.77
Solvent	28.87
R-factor (%)	15.47
R-free (%)	16.76

**Table 2.2.** Processing and refinement statistics for SAP-NADPro crystal, NAD3. Figures in parenthesis denote results for the highest resolution shell (1.58-1.50 Å).

### 2.2.6. Molecular replacement and the Phase Problem

Equation 2.1 demonstrates how all atoms within the unit cell contribute to every spot on the diffraction pattern. A structure factor for each reflection observed on the diffraction pattern can be calculated by Fourier summation of the diffractive contribution of all

atoms in the unit cell. As each reflection can be viewed as diffraction of x-rays by a particular set of parallel planes in the crystal, only the atoms that lie on that set of planes with contribute constructively to the signal recorded by the detector. Viewing each structure factor as a complex wave, the structure factors from all reflections can be combined and Fourier transform used to construct a map of the electron density that produced the observed diffraction pattern. Performing the Fourier transform is analogous to a lens focusing visible light in a standard microscope.

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx + ky + lz - \alpha'_{hkl})} \quad \text{Equation 2.5.}$$

The Fourier transform in Equation 2.5 shows the method for calculating the electron density,  $\rho(x, y, z)$  at coordinates  $x, y, z$ , where  $V$  is the volume of the electron density,  $|F_{hkl}|$  is the amplitude of a structure factor with Miller indices  $h, k, l$  and phase  $\alpha'_{hkl}$ . As previously mentioned the amplitude can be calculated from the intensity measured by the detector, but the phase must be obtained by other means.

There are a variety of methods for solving the phase problem. Molecular replacement is one such method, applicable to proteins whose structures have already been solved by x-ray crystallography and can be downloaded from the Protein Data Bank. In molecular replacement, the coordinates of a previously determined structure are used to calculate new initial phases for the new experimental data. The coordinates of the model structure are translated into the new asymmetric unit and new phases calculated based on the position and orientation of the protein. The calculated phases are then input with the observed amplitudes into calculated structure factors,  $F_{\text{calc}}$ , and the results compared to the values for the experimentally observed structure factors,  $F_{\text{obs}}$ . Phases are calculated for many different orientations of the protein in the asymmetric unit; the result that gives the smallest difference between the calculated and observed data is selected as the best solution and used as the initial phase estimates for structure refinement. In the case of human SAP, the structure has already been solved by protein crystallography (Emsley *et al.*, 1994). Coordinates for the SAP crystal structure 1SAC were downloaded from the PDB. The coordinate file was edited to remove all cations and ligands; leaving only the SAP pentamer as the search model. Firstly, the number of molecules in the

asymmetric unit was estimated based on the cell volume and the molecular weight of the protein, the CCP4 program Matthews determined that one SAP pentamer was present with a probability of 99% (Matthews, 1968). Based on this estimation the molecular replacement program, Phaser (McCoy *et al.*, 2007) was used to search for a single copy of SAP in the asymmetric unit.

Phaser utilises Maximum Likelihood (ML) algorithms to determine new coordinates for the search model within the unit cell. ML is a branch of statistics that calculates the likelihood of a model given the data defined as the probability of the data given the model (McCoy *et al.*, 2007). For molecular replacement this means calculating phases based on the coordinates of the model within the asymmetric unit and determining the probability of observing those phases based on the data. This is achieved by first periodic rotations of the search model within the asymmetric unit and then translating the correctly orientated model to give a set of coordinates most likely to produce the observed data. The rotation and translation searches are performed separately because it dramatically reduces the computational load of molecular replacement to a level that can be handled by desktop computers in a reasonable period of time.

The most probable molecular replacement solution for the NADPro-SAP data was written into a PDB file containing the new coordinates and a new MTZ file containing the initial phase estimates. Phaser calculates log-likelihood gain (LLG) values to determine how much better the new coordinates for the search model represent the observed data compared to a random placement of atoms in the asymmetric unit (McCoy *et al.*, 2007). A Z-score is then computed to show by how many standard deviations the LLG exceeds the average LLG of a set of randomly arranged atoms. A translation function Z-score (TFZ) of greater than 8.0 indicates a certain molecular replacement solution. For NADPro-SAP the molecular replacement solution output by Phaser had a LLG of 14690.3 and this corresponded to a TFZ of 72.9.

The initial phases contain significant errors as they mostly represent the model structure used for molecular replacement. Repeating rounds of refinement of atomic coordinates and calculation of improved phases are employed to construct a model that best represents the electron density that produced the observed diffraction pattern.

### 2.2.7. Refinement

The output molecular replacement solution is a PDB file with the coordinates of the search model placed in the new unit cell. The coordinates from the model allow calculation of phases, which when input into the structure factor equation allow generation of an electron density map. Errors between the calculated phases and the true phases of the experimental data are minimised through refinement. Refinement is performed through ‘reciprocal-space’ phase improvements and ‘real-space’ coordinate refinement. Refinement was performed using Refmac, CCP4 (Murshudov, 1997), while model building was performed using Coot (Emsley and Cowtan, 2004). Refmac utilises log-likelihood calculations to achieve a structure solution with the highest probability given the data. Log-likelihood uses Bayesian statistics that weigh experimental observations against prior information. Prior information includes biochemical knowledge about bond lengths, angles and non-crystallographic symmetry (NCS). Inclusion of prior knowledge is essential for protein crystallography because the data-to-parameter ratio is low.

Model parameters, such as x, y, z coordinates, can be refined based on experimental data, while the prior knowledge restraints, aims to prevent the model containing unrealistic chemical or geometric properties. SAP is a pentameric protein composed of 5 identical subunits arranged radially, which means a 5-fold rotation axis runs through the centre of the pentamer. No crystal system has a 5-fold symmetry element, therefore the axis is non-crystallographic and can be used as a restraint in refinement. The subunits must have identical folds, but may differ in small elements such as the conformation of side chains and flexible loop regions. At the beginning of refinement tight NCS restraints were applied to improve the data-to-parameter ratio of the data when the phase error was at its highest, as refinement proceeded the NCS restraints were loosened to allow local deviations in the structure to be borne out.

The output PDB file from MR was input into Refmac with the reflection MTZ file and refined as a rigid body to improve the initial placement of the pentamer in the asymmetric unit; this gave a starting R-work value of 33.64% and a R-free of 33.25%. The R-work is a measure of the agreement between the observed diffraction data and the diffraction data that would be expected based on the model (Brünger, 1992).

Brünger introduced the R-free value as a cross-validation statistic to highlight potential bias in the R-work caused by over-refinement. Prior to refinement, a small proportion of the experimental data (5% or 2500 reflections, whichever is lower) is selected as the R-free set and omitted from use in refinement. As the R-free data is not used to calculate the phases for the electron density map it can be used as a less biased judge of the model's quality.

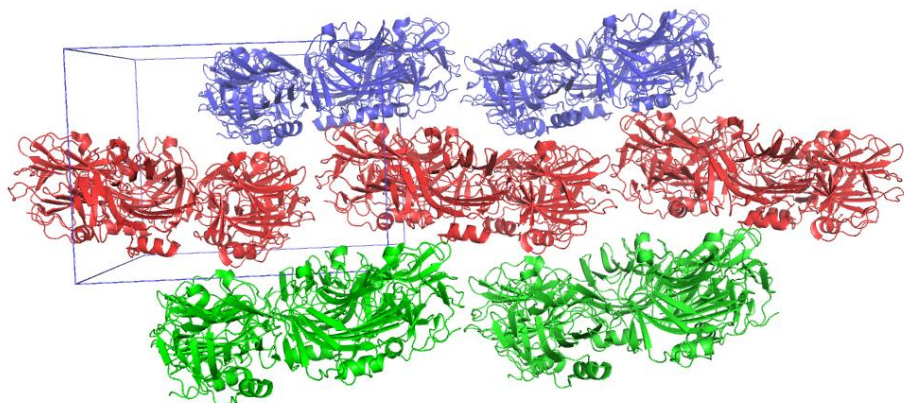
The pentamer model and electron density maps were then observed as  $\sigma A$  weighted 2Fo-Fc and Fo-Fc maps (Read, 1986) using the molecular graphics program Coot (Emsley, 2004). Fo ( $F_{obs}$ ) are the observed structure factors from the reflection, MTZ, file, while Fc ( $F_{calc}$ ) are the calculated structure factors based on the model. Molecular graphics programs, such as Coot, can display maps that highlight the differences and similarities between Fo and Fc to aid model building. The map showed 2 large spheres of difference density around Asp-138 for each protomer, this region is known to be the calcium binding pocket of SAP (Emsley *et al.*, 1994). Therefore, 2 calcium ions were built into the spheres of difference density for each protomer, 10 in total.

The subsequent rounds of alternating reciprocal-space restrained refinement and real-space model building involved placing N-acetylglucosamine at Asn-32 of each subunit as previously reported (Tennent and Pepys, 1994), rebuilding flexible loop regions to trace electron density and the placement of solvent, water and glycerol molecules. The final R-work value for the refined model was 15.47% and 16.76% for the R-free. Root-mean-square (RMS) values for bond lengths and angles are a simple method for determining deviation of the model from biochemical ideal bond geometry; for the NADPro-SAP structure, the final root-mean-square RMS Bond value was 0.007Å and the RMS Angle was 1.26°.

## 2.3. Discussion

### 2.3.1. Crystal Packing

The molecular graphics program PyMol (PyMol molecular graphics system, Version 1.3 Schrödinger LLC) was used to demonstrate the packing of SAP pentamers in the crystal (Figure 2.11). The pentamers are stacked A-face to B-face in layers that are staggered by displacement of half a pentamer's width from the layer beneath both along and perpendicular to the plane of the paper. There are no large gaps between the pentamers indicating that molecular replacement found all the copies of SAP in the asymmetric unit.



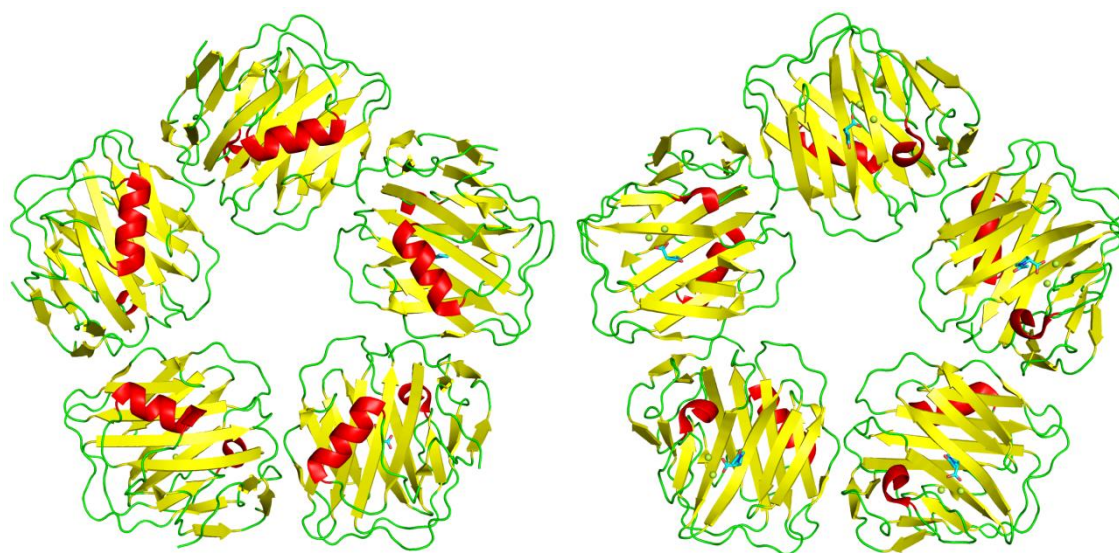
**Figure 2.11.** Crystal packing of SAP in crystal NAD3 with B-face of pentamer uppermost in all cases. Pentamers are coloured to clarify the different layers of protein molecules within the crystal. The unit cell is depicted by the blue outline.

### 2.3.2. Protein Structure

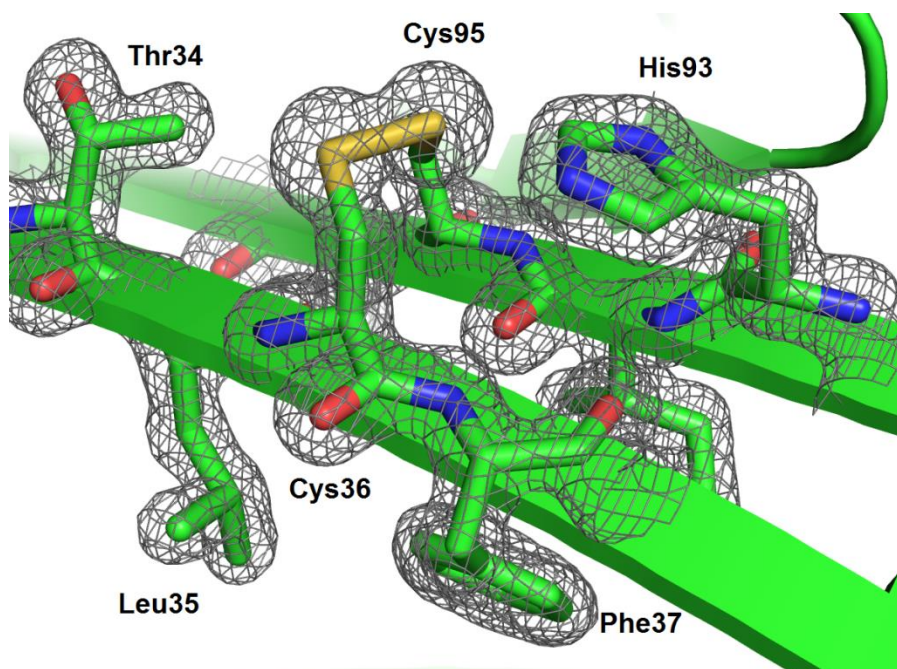
The refined SAP structure in this investigation conforms to those previously reported (Emsley *et al.*, 1994; Srinivasan *et al.*, 1994). The refined structure shows 5 identical protein chains arranged radially (Figure 2.12). Each protomer chain is folded into a flattened  $\beta$ -jellyroll topology. The  $\beta$ -jellyroll is formed by two anti-parallel  $\beta$ -sheets each 7 strands in length (Figure 2.13). In addition there is a short  $\alpha$ -helix on the A-face of each pentamer and a calcium binding loop on the opposing B-face. These observations provide certainty that the desired protein was crystallised and that the



Chapter 2. X-ray crystal structure of Serum Amyloid P component in complex with N-acetyl D-proline molecular replacement solution is correct. In order to validate this observation, a number of tools and calculations were employed to assess the model geometry and the consistency of the model with the observed electron density.



**Figure 2.12.** A and B faces of NADPro-SAP structure in crystal NAD3.  $\alpha$ -helix (red),  $\beta$ -sheet (yellow) and loops/flexible regions (green).



**Figure 2.13.** Example of 2Fo-Fc electron density map (grey mesh,  $1.8\sigma$ ) generated from the NADPro-SAP dataset and the protein model built within it as a ball and stick and cartoon representation. Pictured is part of two neighbouring  $\beta$ -strands containing cysteines that form the single disulphide bond in each SAP subunit. Model colours:  $\beta$ -sheet cartoon (green ribbon); ball and stick diagram: carbon (green), oxygen (red), nitrogen (blue) and sulphur (yellow).

### 2.3.3. Structure Validation

#### 2.3.3.1. Protein Geometry

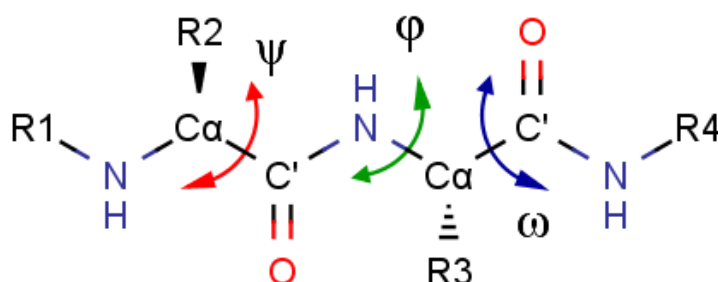
The bond lengths and angles are restrained during refinement, but some deviations from ideal geometry are expected due to steric strain imposed on bonds, protein folding and electrostatic interactions with neighbouring functional groups. Conversely, the low data-to-parameter ratio of x-ray crystallographic data can lead to over-refined structures where unrealistic or impossible geometry is imposed upon the model by phase errors that manifest themselves as erroneous peaks in the electron density map. It is important to strike a balance between restraining and over-refining a structure and this can be achieved through monitoring R-factors and the aggregate RMS deviations from ideal bond lengths and angles.

The difference between the R-work and the R-free is used to assess model bias. A general rule is that the R-free should be no greater than 5% higher than the R-work. In the case of the NADPro-SAP structure the final R-work value was 15.47% with an R-free of 16.76%, this provides confidence that the model is not significantly biased due to over-refinement.

RMS values for bond lengths and angles aggregated across the whole structure can be compared with other solved structures in the PDB. The acceptable upper limit of RMS values for bond lengths and angles is generally considered to be 0.02Å and 2.00° respectively, although there is no definite agreement on these values amongst crystallographers. The NADPro-SAP structure presented here is of reasonably high resolution, so there is more crystallographic data to influence the model by providing experimental observations for deviations from ideal geometry. Therefore with RMS bond length and angles of 0.007Å and 1.26° respectively, the overall bond geometry of the NADPro-SAP structure can be considered reliable.

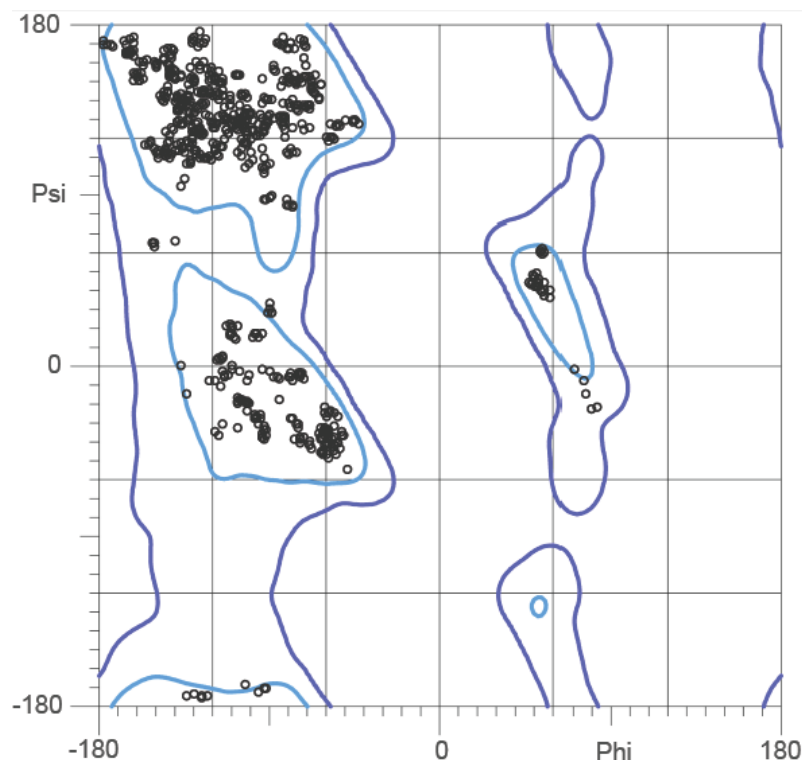
### 2.3.3.2. Ramachandran Plot

The Ramachandran plot is a useful tool for assessing the geometry of the protein backbone (Ramachandran *et al.*, 1963). The Ramachandran plot is a graph of the  $\phi$ - $\psi$  torsion angles between the N-C $_{\alpha}$  and the C $_{\alpha}$ -C' bonds of a peptide (see Figure 2.13). Ramachandran calculations are not used for refinement in Refmac so can be applied to unbiased validation of model geometry.



**Figure 2.14.** Backbone torsion angles of a polypeptide chain.

Ramachandran *et al.* (1963) observed that combinations of backbone torsion  $\phi$  and  $\psi$  angles coalesced around a range of allowed values depending on the identity of the amino acid residue and its position within a polypeptide sequence. Plotting backbone torsion angles on a contour map reveals a number of forbidden regions which would bring atoms into too close contact. For ideal geometry all backbone torsion angles should be within the favoured regions of the plot. The Ramachandran plot for the general case of NADPro-SAP structure is pictured in Figure 2.15. The plot shows all residues in the allowed region of the plot and the vast majority within the favoured region. This shows that the protein backbone does not contain unrealistic geometry and is therefore not over-refined, nor does it contain areas where the main protein chain is under excessive strain.



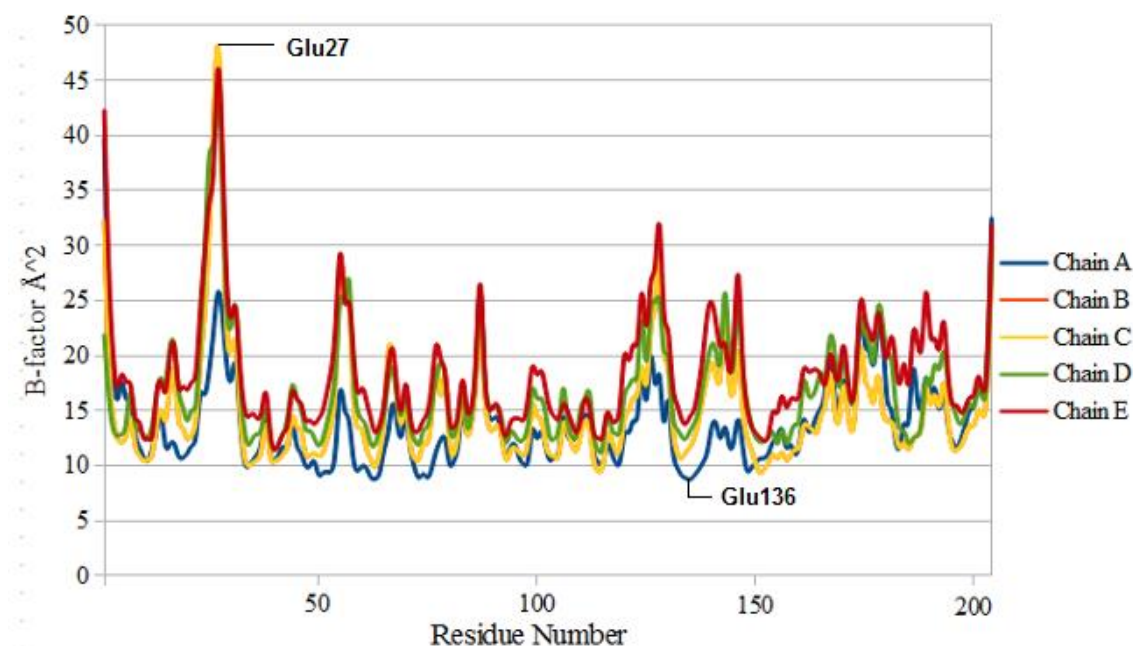
**Figure 2.15.** Ramachandran plot for the general case of NADPro-SAP protein backbone torsion angles generated using the Molprobity server for all subunits (Chen *et al.*, 2010). Light blue contour indicates favoured regions, dark blue contour indicates allowed regions. The Phi and Psi angles shown on each axis relate to the backbone torsion angles depicted in Figure 2.14.

#### 2.3.3.3. B-factor analysis

The atomic displacement parameter or B-factor is an atomic scattering factor modifier that is used to account for displacement of electron density from its mean position due to factors such as greater thermal motion. Higher thermal motion is expected in flexible regions of a protein such as loops or disordered side chains. The average B-factor for a protein model tends to be proportional to the resolution of the data. Reciprocal space refinement programs, such as Refmac, increase the B-factor to compensate for areas of poor electron density in the map. This reduces the contribution from inaccurate phases and is crucial in minimising the error in the overall calculated phase.

B-factors for particular residues or groups of residues can be compared to the average B-factor to identify problematic regions where the data is poorly described by the model. The cause of poor electron density can be high flexibility of residues or it could be a sign that parts of the model are built incorrectly. The average B-factor for the

protein atoms in the NADPro-SAP structure was 14.72. Figure 2.16 demonstrates the B-factor per residue for each chain of the refined protein structure. Although the B-factors fluctuate significantly throughout each chain, there is no region that is systematically worse than the rest of the chain. The biggest deviations from the average B-factor are observed for residues 25-28, 55-57 and 122-130, which all correspond to flexible loop regions of SAP.

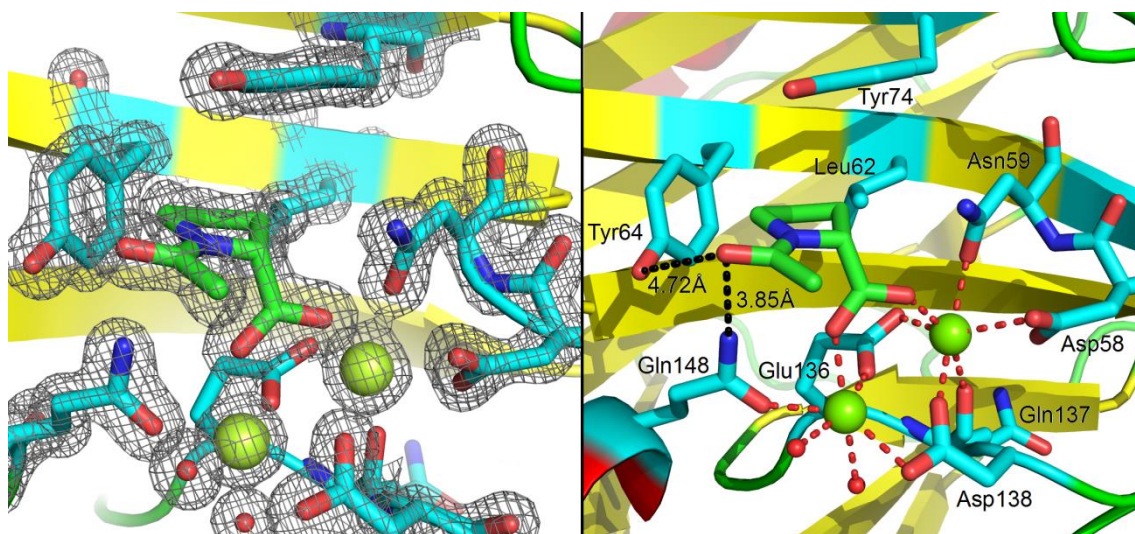


**Figure 2.16.** B-factor per residue of each chain for NADPro-SAP structure. There is clear difference in the B-factors between Glu27, which is part of flexible loop region and Glu136, which is a crucial residue for metal binding.

#### 2.3.3.4. Binding site analysis

The refined NADPro-SAP structure shows, each calcium ion is bound by a loop region on the B-face of SAP. The first calcium ion is bound by the side chain functional groups of Asp58, Asn59, Glu136 and Asp138, the main chain carbonyl of Gln137 with the carboxyl group of NADPro completing the coordination sphere of the metal ion (Figure 2.17). The second calcium ion is more loosely bound with the protein coordinating the metal through the side chains of Glu136, Asp138 and Gln148. The rest of the coordination sphere is completed by two water molecules and the carboxyl group of the ligand. These observations are consistent with those made by Emsley *et al.* (1994).



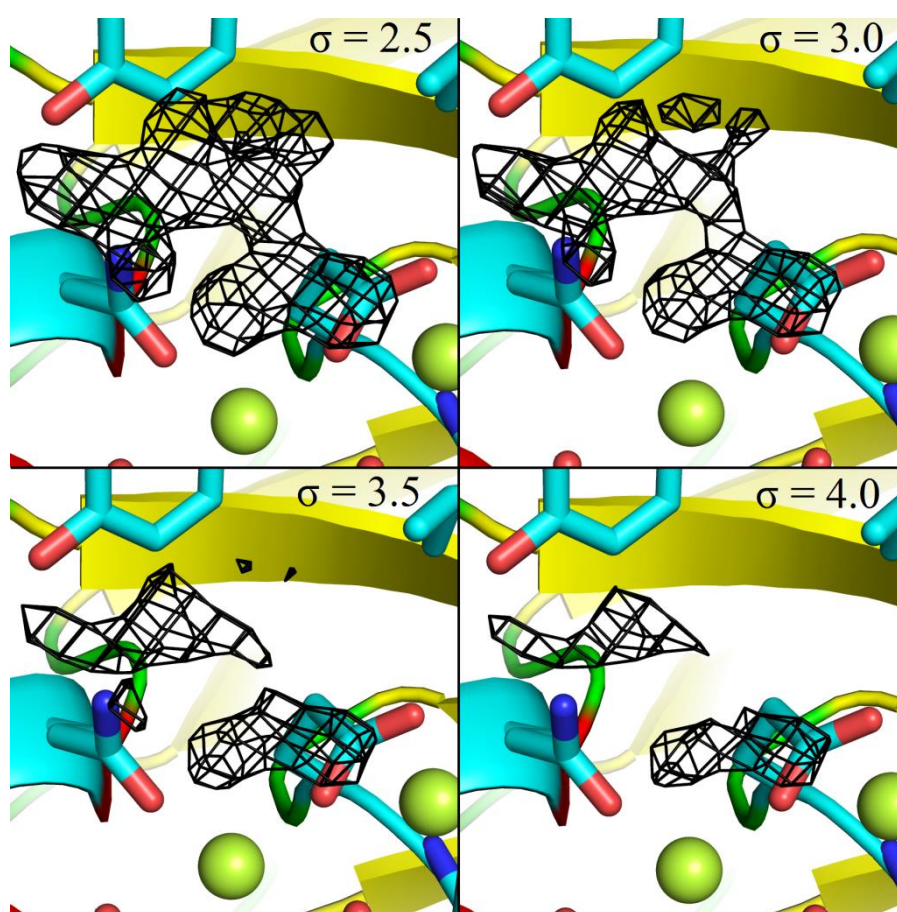


**Figure 2.17.** Ball and stick representation of NADPro bound into the double calcium binding site of SAP. Left: 2Fo-Fc electron density map (grey mesh) of NADPro-SAP data drawn at 1.8 $\sigma$ . Clear density is seen for NADPro within the binding site with a T-shaped protrusion pointing towards the two calcium ions (limon yellow spheres) and the pyrrolidine ring fitting into the adjacent hydrophobic pocket. Right: image with red dotted lines show coordination of calcium ions. Black dotted lines show distances between carbonyl oxygen of acetyl of NADPro and Gln148 and Tyr64. SAP atom colours: carbon (turquoise), oxygen (red) and nitrogen (blue), red spheres (water). SAP Cartoon:  $\beta$ -sheet (yellow),  $\alpha$ -helix (red) and loop (green); NADPro colours: carbon (green), oxygen (red) and nitrogen (blue).

Prior to adding the ligand to the model there was a clear area of difference density for NADPro around the calcium binding site of each protomer that showed a ring structure and two T-shaped protrusions that were attributed to the acetyl and carboxylate of the ligand (Figure 2.17). Coordinates and restraints for the NADPro ligand were built using the Prodrgr server (Schüttelkopf and van Aalten, 2004). NADPro coordinates the calcium ions through its carboxylate with the pyrrolidine ring fitting into the hydrophobic pocket (Figure 2.17). This arrangement is in agreement with previously reported binding of ligands with SAP (Purvis Thesis, 2002; Pepys *et al.*, 2002).

Despite the high resolution of the structure, the peaks in the electron density map for the ligand did not provide conclusive evidence over the proper orientation of the acetyl group of NADPro. Comparing the nature of the atoms in the acetyl group, oxygen has 8 electrons and an electronegativity of 3.44 on the Pauling Scale (Housecroft and Constable, 2006), carbon has 6 electrons and an electronegativity of 2.55. Therefore, we would expect the carbonyl oxygen of the acetyl group to demonstrate a slightly higher peak in the electron density map than the carbon of the methyl group if there is a fixed

conformation. The calcium binding site of each subunit was observed separately and the electron density map redrawn at different standard deviation ( $\sigma$ ) values to find the strongest areas of electron density for the acetyl group of NADPro. This method proved problematic as the differences between the intensity of the electron density map for each part of the acetyl group were minimal. For this reason, the process was repeated after calculation of an NCS difference map in Coot to show the average electron density distribution across the ligand (Figure 2.18). This showed unequivocally a larger peak for density *trans* to the Ca of NADPro, which supports the notion that the acetyl group favours the *cis* conformation when averaged across all subunits.



**Figure 2.18.** Omit map of N-acetyl D-proline electron density calculated at sigma values between 2.5-4.0. A central ring with two T-shaped protrusions is clearly discernible when  $\sigma = 2.5$ . While most electron density has been lost by  $\sigma = 3.5$ , there is still clear electron density for the half of the acetyl group that is *trans* to the carboxyl group. SAP atom colours: carbon (turquoise), oxygen (red) and nitrogen (blue), calcium ions (lemon green spheres) water (red spheres). SAP Cartoon:  $\beta$ -sheet (yellow),  $\alpha$ -helix (red) and loop (green).

Beyond direct observation of the electron density map of the ligand, the refined

structure shows a number of amino acid residues in close proximity to the NADPro binding site that could influence the conformation of the ligand through electrostatic interactions. Prior to this investigation the phenyl side chain of Tyr64 was identified as a potential hydrogen bond partner for the carbonyl groups of NADPro and CPHPC.

Hydrogen bonds form between a donor, which is an electronegative species with an attached partially-positively charged hydrogen that it can “donate”, and a species with a free lone pair of electrons capable of accepting the hydrogen. For hydrogen bond formation, the acceptor and donor species must be separated by a distance of between 2.2-4.0 Å (Jeffrey, 1997). The strength of the interaction diminishes rapidly as the distance between the acceptor and donor increases. Using Coot, the average distance between the phenyl oxygen of Tyr64 and the carbonyl oxygen of the NADPro was measured to be 4.72 Å. Therefore, no hydrogen bond can exist between these two functional groups (Figure 2.17). However, it was observed that the side chain amide of Gln148 is an average distance of 3.85Å from the acetyl group of the ligand, which is close enough for a weak electrostatic interaction.

### **2.3.5. N-acetyl D-proline as a headgroup for multivalent binding**

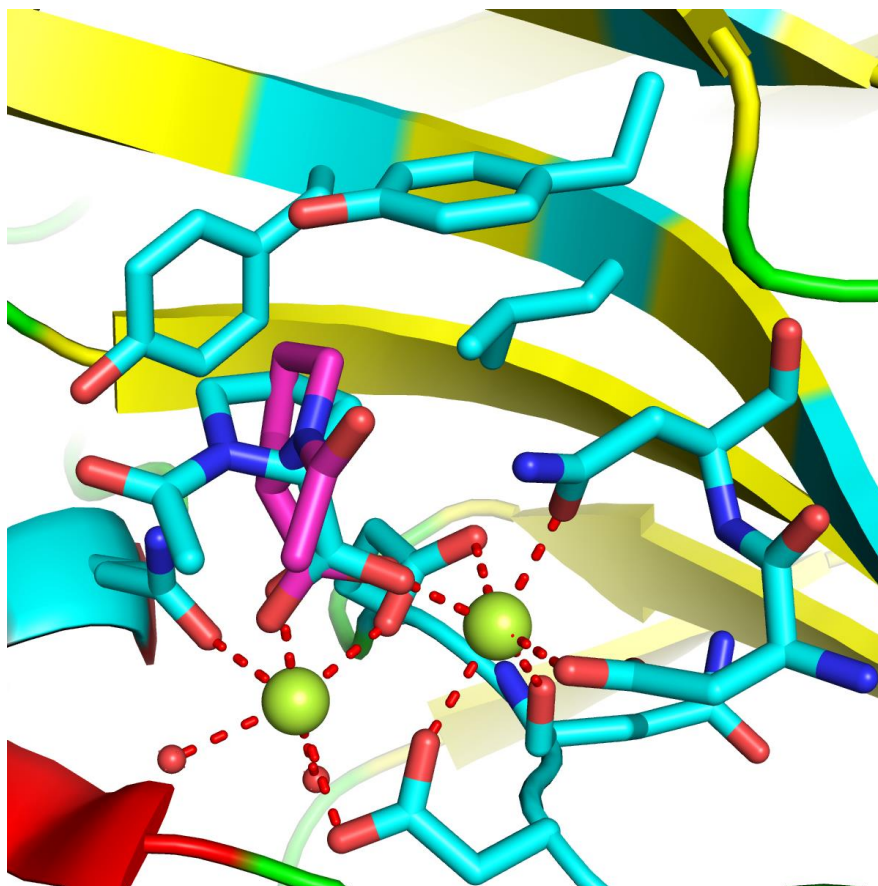
The crystal structure presented here demonstrates how N-acetyl D-proline binds to SAP when unencumbered by any other functional groups that could induce strain into protein-ligand interaction preventing the ligand from binding in the ideal conformation. These results also provide valuable information for designing multivalent ligands for SAP, in particular the choice of headgroup and the design of the linker used to ligate it to the rest of the molecule. The ideal multivalent ligand should preserve the binding geometry of the univalent headgroup, so it is important that the linker used to combine the 5 monomeric D-prolines into a single multivalent ligand is simple and does not contain bulky functional groups that could sterically hinder binding. Some weak hydrogen bonding character may be present between Gln148 and the acetyl group of NADPro. This may be retained in a multivalent ligand if the D-proline is attached to the rest of the construct through a peptide bond. However, the bond, if it does exist, is likely to make only a minor contribution to the binding affinity compared to the contribution from the di-metal coordination of NADPro's carboxyl group. Therefore, it is not imperative to make special considerations for a Gln148 hydrogen bonding group in a



Chapter 2. X-ray crystal structure of Serum Amyloid P component in complex with N-acetyl D-proline

multivalent ligand for SAP. The design need only ensure that the terminal residue is a D-proline presented with a free carboxyl group and the freedom to bind in its preferred conformation.

All known ligands for SAP bind in calcium-dependent manner through either a carboxyl or phosphate group and N-acetyl D-proline is no exception (Emsley *et al.*, 1994; Hohenester *et al.*, 1997; Pepys *et al.*, 2002; Purvis thesis, 2002; Mikolajek *et al.*, 2011; Kolstoe *et al.*, 2014). But the importance of the hydrophobic interaction between the pyrrolidine ring of NADPro and the side chains of Leu62, Tyr64 and Tyr74 should not be overlooked. Kolstoe measured the binding affinity for both D and L proline with SAP by Isothermal Titration Calorimetry, which showed D-proline to have a dissociation constant of 18.6 $\mu$ M, almost 20-fold stronger than the 322 $\mu$ M dissociation constant for L-proline (Kolstoe thesis, 2005). The high resolution crystal structure of NADPro-SAP presented here shows the key reason for this difference, the pyrrolidine ring interaction with the hydrophobic pocket of SAP. The structure of SAP in complex with N-acetyl L-Proline (NALPro) downloaded from the PDB (PDB code: 4AVS) shows the pyrrolidine ring pointing away from the hydrophobic pocket (Figure 2.19), which is no doubt the cause of the reduced binding affinity. The tight fit of D-proline into the binding site of SAP provides encouragement that D-proline is the ideal headgroup to present to SAP as part of a multivalent ligand and for this reason it was used in the synthesis of multivalent ligands that will be presented in Chapters 4 and 5.



**Figure 2.19.** Calcium-dependent binding of D-proline (turquoise carbon atoms; PDB file: 4AYU) and L-proline (magenta carbon atoms; PDB file: 4AVS) with SAP.

## Chapter 3. X-ray crystallographic structure of rat C-reactive Protein in complex with bivalent ligand BPC8

### 3.1. Introduction

The exact function of the Pentraxins' in humans is not known, but SAP and C-reactive Protein (CRP) have become key targets in the development of therapeutics for amyloidosis (Pepys *et al.*, 2002) and cardiovascular disease (Pepys *et al.*, 2006), respectively. Elucidating the functions of the Pentraxins in humans is hampered by the fact that no deficiency or polymorphism of either protein has been reported, preventing an 'experiment of nature' to decipher their function (Pepys *et al.*, 1997). However, investigations of structural differences in related proteins from other organisms can improve understanding of human pentraxins, particularly with regard to ligand binding specificity.



**Figure 3.1.** The structure of phosphoethanolamine (PE; left) and phosphocholine (PC; right).

The archetype ligand for human SAP (hSAP) is phosphoethanolamine, for CRP it is phosphocholine (Figure 3.1; Mikolajek *et al.*, 2011). hCRP also binds to phosphoethanolamine, but hSAP shows no interaction with phosphocholine. In contrast, rat SAP (rSAP) has been observed in complex with phosphocholine by x-ray crystallography (Mikolajek thesis, 2008). The structure of the calcium-binding site is conserved between human and rat SAP, except for a single conservative substitution of Glu66 to Asp66 in rats. This single change moves a hydrogen bonded water molecule further from the calcium-binding site of SAP. This has been identified as a key residue in preventing phosphocholine binding by human SAP, thus demonstrating the value of investigating related proteins from animals.

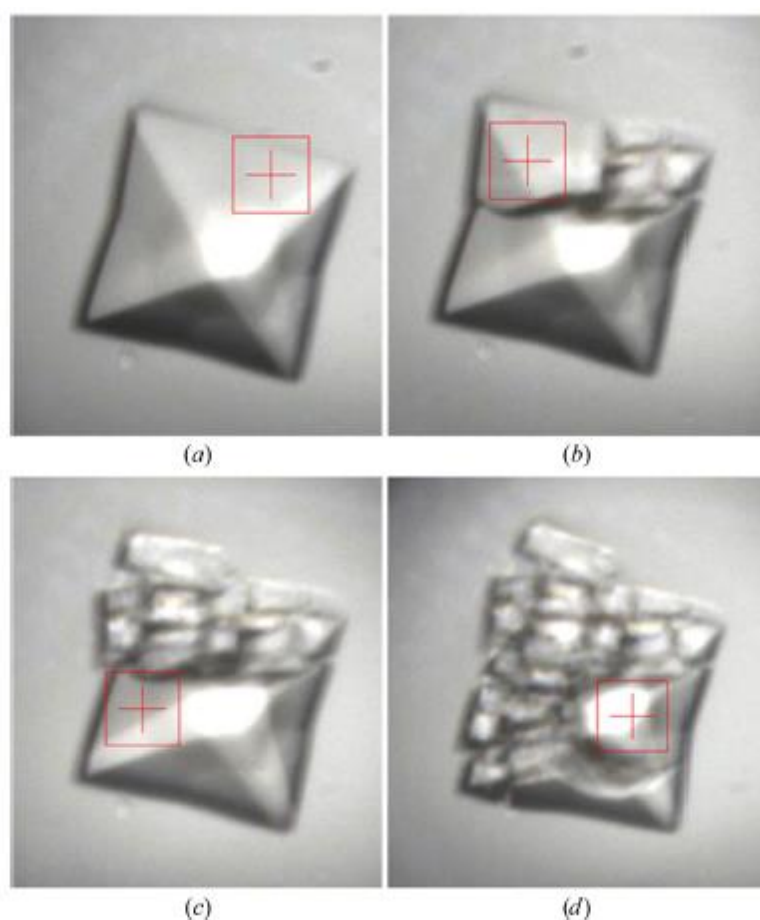
The other rat pentraxin, rat CRP also has a number of structural features that make its crystal structure worthy of investigation. rCRP is glycosylated by a single biantennary

oligosaccharide, which is common amongst SAP, but highly unusual for mammalian CRP (De Beer *et al.*, 1982; Sambasivam *et al.*, 1993). rCRP also has an extended C-terminal amino acid sequence of LTDCCES (Rassouli *et al.*, 1992). It has also been observed that two subunits per pentamer form an intermolecular disulphide bond between their extended C-termini, with the intramolecular disulphide bonds forming between the neighbouring cysteines on the remaining subunits. Intramolecular disulphide bonds formed between neighbouring cysteines are rare, but is known to occur in methanol dehydrogenase from *Methylococcus capsulatus* (Culpepper and Rosenzweig, 2014)

As a result, numerous attempts have been made to crystallise rat CRP (rCRP; Hopkins *et al.*, 1994; Mikolajek thesis, 2008). Despite the fact that claims of solving the x-ray crystal structure of rCRP have been made, the results remain unpublished (Shrive *et al.*, 1996). Difficulties in crystallising rCRP both ligand-free and in complex centre around a high degree of radiation sensitivity of crystals produced (Hopkins *et al.*, 1994; Mikolajek thesis, 2008); to the point that diffraction is completely lost after just a few frames of data have been collected. Radiation damage from the x-ray beam afflicts all protein crystals, particularly when conducting diffraction experiments on modern synchrotrons which produce high energy x-rays (Ravelli *et al.*, 2006).

Radiation damage in crystallography is caused by absorption of photons by either the photoelectric effect (total absorption of the incident radiation leading to ejection of an inner shell electron) or Compton scattering (inelastic scattering of a photon, which can also cause emission of an electron; Garman, 2010). Inner shell electrons liberated by x-ray absorption are able to travel along the protein backbone by quantum tunnelling and will localise at sites of high electron affinity. The mobility of liberated electrons means some amino acids show greater susceptibility to radiation damage and leads to local damage occurring in a reasonably predictable order. Local damage affects individual regions and bonds of a protein and occurs first through elongation and then breakage of disulphide bonds followed by decarboxylation of aspartate and glutamate residues and then breakage of C-S bonds of methionine side chains (Weik *et al.*, 2000; Burmeister, 2000). In contrast, global damage is a reduction of crystalline order caused by formation

of hydrogen bubbles in the crystal and with increasing x-ray exposure it manifests itself through changes in unit cell volume and mosaicity and loss of diffraction intensity (Teng and Moffat, 2000; Ravelli *et al.*, 2002; Burmeister, 2000; Ravelli and McSweeney, 2000; Weik *et al.*, 2000; Müller *et al.*, 2002; Meents, 2009; Meents, 2010). The loss of periodicity is graphically demonstrated by images of crystals of bovine enterovirus 2 disintegrating during room temperature diffraction experiments reported by Axford *et al.* (2012; Figure 3.2)



**Figure 3.2.** Crystal of bovine enterovirus 2. a) prior to exposure, b) after 0.5s exposure to upper right quadrant, c) after 0.5s exposure to upper left quadrant, d) after 0.5s exposure to lower left quadrant. Red cross-hair denotes position of x-ray beam for next exposure. Adapted from Axford *et al.* (2012).

The extent of radiation damage is greatly mitigated by cryo-cooling crystals and performing diffraction experiments at 100K (Garman, 1999), but damage will still occur and some crystals will show greater susceptibility depending on a number of factors,

such as the nature and packing of the protein. It has been shown that global loss of diffraction intensity is caused principally by formation of hydrogen bubbles within the unit cell, rather than through damage to residues involved in intermolecular contacts between proteins that facilitate the crystal packing (Meents, 2009; Meents, 2010). However, the exact crystal packing is still important and it is likely that increasing the number or strength of intermolecular contacts could produce crystals with longer diffraction lifetimes. Therefore, if cryo-conditions cannot sufficiently extend the diffraction lifetime of a crystal, then alternative crystallisation conditions should be employed to attempt to crystallise the protein in an alternative spacegroup with different crystal packing. This may require the addition of agents to crystallisation cocktails that are known to improve crystal resilience; divalent cations such as cadmium have been reported to be particularly effective additives (Trakhanov *et al.*, 1998). However, producing diffraction quality crystals may still prove intractable and it may be necessary to return to sparse matrix screens to begin a *de novo* search for new crystallisation conditions. Ultimately, even successfully growing crystals in different crystallisation conditions may not improve diffraction lifetime if the crystal packing of the protein is inherently unstable.

Co-crystallisation of bivalent ligands, CPHPC and dAMP with SAP produces crystals with unique packing arrangements that have not been observed in crystals of SAP in complex with univalent ligands. In theory, bivalent ligand, BPC8, which binds to human CRP through the same non-covalent cross-linking mechanism as CPHPC binding to SAP, should produce crystals of rCRP with unique crystal packing. A new crystal packing arrangement could result in different inter-protein contacts that overcome the radiation sensitivity observed by Hopkins *et al.* (1994) and Mikolajek (thesis, 2008). BPC8 should bridge strong non-covalent interactions between pentamers adding new intermolecular contacts that could increase stability of the crystal structure. This section describes the results of the crystallisation and the potential to use multivalency to solve crystallographic problems.

## 3.2. Materials and Methods

### 3.2.1. Protein purification

Rat CRP was isolated from normal rat serum by calcium dependent affinity chromatography with phosphoethanolamine covalently immobilized to carboxethyl Sepharose. After extensive washing with calcium-containing TC buffer (10mM Tris/HCl, 140mM NaCl, 5mM CaCl<sub>2</sub>, pH 8.0) rCRP was eluted with 1mM phosphocholine in the same buffer. A second cycle of calcium-dependent binding to PC-Sepharose was employed to concentrate the protein, which was finally eluted in TE buffer. rCRP was dialysed against TC buffer. Purity was checked by 8-18% SDS polyacrylamide gradient gel electrophoresis (purity > 99%).

Bis-phosphocholine ligand, BPC8, was a gift from Prof. Pepys laboratory (London, UK). BPC8 is a bivalent molecule where two phosphocholine molecules are cross-linked through a phosphate oxygen by an 8 carbon alkyl chain (Figure 3.3).



**Figure 3.3.** bis-Phosphocholine ligand for CRP, BPC8.

### 3.2.2. X-ray Crystallography

Crystals of rCRP were grown by hanging drop vapour diffusion (McPherson, 1982) following extensive screening with different crystallisation cocktails containing varying excesses of BPC8. Harvested crystals were flash-cooled to 100K by plunging into liquid nitrogen using a nylon loop with the crystal suspended in a small quantity of mother liquor without additional cryoprotectant. rCRP diffraction data was collected on beamline I04 at Diamond Light Source (Didcot, UK). Data was processed using Mosflm (Powell, 2006) and programs from the Collaborative Computer Project 4 suite

(CCP4, 1994). Molecular replacement was performed using Phaser (McCoy *et al.*, 2007) with PDB coordinates 1B09 (Thompson *et al.*, 1999). The rCRP-BPC8 complex was refined using Phenix (Adams *et al.*, 2010). Model building was performed with Coot (Emsley and Cowtan, 2004). Coordinates for BPC8 were generated using the ProdrG server (Schüttelkopf and van Aalten, 2004), geometry restraints for the ligand were generated using Phenix.elbow.

### **3.3. Results**

#### **3.3.1. Protein crystallisation**

Crystals of rCRP were obtained from a stock solution of 6.1mg ml<sup>-1</sup> protein in TC buffer by cocrystallisation with a 10-fold molar excess of BPC8 at 4°C. The protein-ligand solution was mixed with an equal volume of reservoir solution containing 150mM acetate buffer pH 5.0, 45% w/v 2-methane-2,4-pentanediol (MPD) and 50mM calcium chloride. Crystals appeared after 4 weeks, growing to maximum dimension of 0.3 x 0.1 x 0.1 mm after 9 weeks. The concentration of MPD was sufficiently high for it to act as a cryoprotectant, so crystals were flash-cooled without addition of further cryoprotectant.

#### **3.3.2. Diffraction data collection and processing**

rCRP crystals were flash cooled in liquid nitrogen at 100K in a small quantity of solvent from the mother liquor prior to data collection. In view of previous reports of extreme radiation sensitivity, data was collected in two passes with low and then high beam attenuation. The two data sets were later merged into a single Mtz file during data processing. The methodology employed for collecting multiple datasets from a single crystal is controversial in the crystallographic community. It is often asserted that the high resolution pass should be performed first as these reflections are the first to be lost due to radiation damage. In light of previous investigations, radiation sensitivity was expected to be the limiting factor for all reflections. Therefore, the low resolution pass was performed first to acquire a complete dataset using just 20% transmission of the x-ray beam. At the end of the first data collection, diffraction spots were still visible towards the edge of the detector, so a second pass was initiated to capture higher



resolution reflections using 100% transmission of the x-ray beam. Autoindexing in Mosflm recommended hexagonal space group P622 with unit cell dimensions of  $a=b=111.50\text{\AA}$ ,  $c=424.74\text{\AA}$  and  $\gamma=120^\circ$  at a maximum resolution of  $3.2\text{\AA}$  and mosaicity of  $0.45^\circ$ . Full dataset statistics are provided in Table 2.



**Figure 3.4.** X-ray diffraction pattern of rCRP-BPC8 crystal during second data collection pass, using 100% transmission of the beam collecting reflections extending to a maximum of  $2.9\text{\AA}$  using a CCD detector. Bottom right: enlarged image of part of the centre diffraction pane showing the beamstop (lighter grey portion) with the centre of the beam within this space. The closely packed diffraction spots indicate at least one long unit cell axis.

The high symmetry spacegroup was considered particularly beneficial as the rCRP crystals were expected to be highly radiation sensitive; the higher the spacegroup symmetry, the more reflections that are symmetry-related. Therefore, a smaller portion of reciprocal space needs to be sampled to obtain a full dataset of all measurable reflections, so x-ray exposure time can be kept to a minimum. In the first pass  $120^\circ$  of

data were collected with a 1° oscillation angle. In the second x-ray exposure of the same crystal, 180° of data were collected using a 1° oscillation angle as a safeguard against incorrect spacegroup determination by Mosflm. The merging and scaling of the two datasets was performed in Scala. The high symmetry of the spacegroup and the combination of two data collection runs gave a multiplicity of 34.5 overall and 35.9 in the highest resolution shell (3.37-3.20Å) and outer shell completeness of 99.6% with an I/σ(I) of 2.7.

### 3.3.3. Molecular replacement

The sequence homology between rat pentraxins and their human homologue is high, 65% for rCRP (Lei *et al.*, 1985; Woo *et al.*, 1985; Table 3.1). Therefore, a previously solved structure for hCRP was used as the search model for molecular replacement of the rCRP data. Coordinate file 1B09 was downloaded from the Protein Data Bank (PDB) and edited to leave only amino acid residues from a single pentamer (Thompson *et al.*, 1999). The Matthews coefficient (Matthews, 1968) for the structures indicated one pentamer per asymmetric unit with a solvent content of 58%. A molecular replacement solution was obtained using Phaser with an LLG of 493 and a TFZ of 11.8. A subsequent round of rigid-body refinement in Refmac gave an initial R-work of 50.5% and an R-free of 52.55%. The R-factor for a collection of atoms placed randomly within the unit cell is 59% (Rupp, 2010), so the starting R-factors were high and this can be mostly likely attributed to the limited resolution of the data and the increased phase error from using the human homologue of the rCRP as the search model. However, the TFZ score of 11.8 indicates a certain molecular replacement solution and provides confidence that the pentamer was placed correctly within the unit cell and that 1B09 was an appropriate search model.

CRP	10	20	30	40	50
Homo Sapiens	QTDMRKAFV	FPKESDTSYV	SLKAPLTKPL	KAFTVCLHFY	TELSSTRGYS
Rattus Norveigicus	HEDMSKQAFV	FPGVSATAYV	SLEAESKKPL	EAFVCLYAH	ADV--RSFS
	60	70	80	90	100
Homo Sapiens	IFSYATKRQD	NEILIFWSKD	IGYSFTVGGG	EILFEVPEVT	VAPVHICTSW
Rattus Norveigicus	IFSYATKTSF	NEILLFWTRG	QGFSIAVGGP	EILFSASEIP	EVPTHICATW
	110	120	130	140	150
Homo Sapiens	ESASGIVEFW	VDGKPRVRKS	LKKGYTVGAE	ASIILGQEQD	SFGGNFEGSQ
Rattus Norveigicus	ESATGIVELW	LDGKPRVRKS	LQKGYIVGTQ	ASIILGQEQD	SYGGGF <del>D</del> ANQ
	160	170	180	190	200
Homo Sapiens	SLVGDIGNVN	MWDFVLSPDE	INTIYLGGPF	SPNVLNWRAL	KYEVQGEVFT
Rattus Norveigicus	SLVGDIG <del>D</del> VN	MWDFVLSPDQ	INAVYVGRVF	SPNVLNWRAL	KYETHGDVFI
	210	220			
Homo Sapiens	KPQLWP				
Rattus Norveigicus	KPQLWPLTDC	CES			

**Table 3.1.** Aligned amino acid sequences of human CRP and rat CRP. Highlighted residues show areas where sequences diverge, with blue representing a conserved change and red a non-conserved sequence difference.

### 3.3.4. Refinement and model building

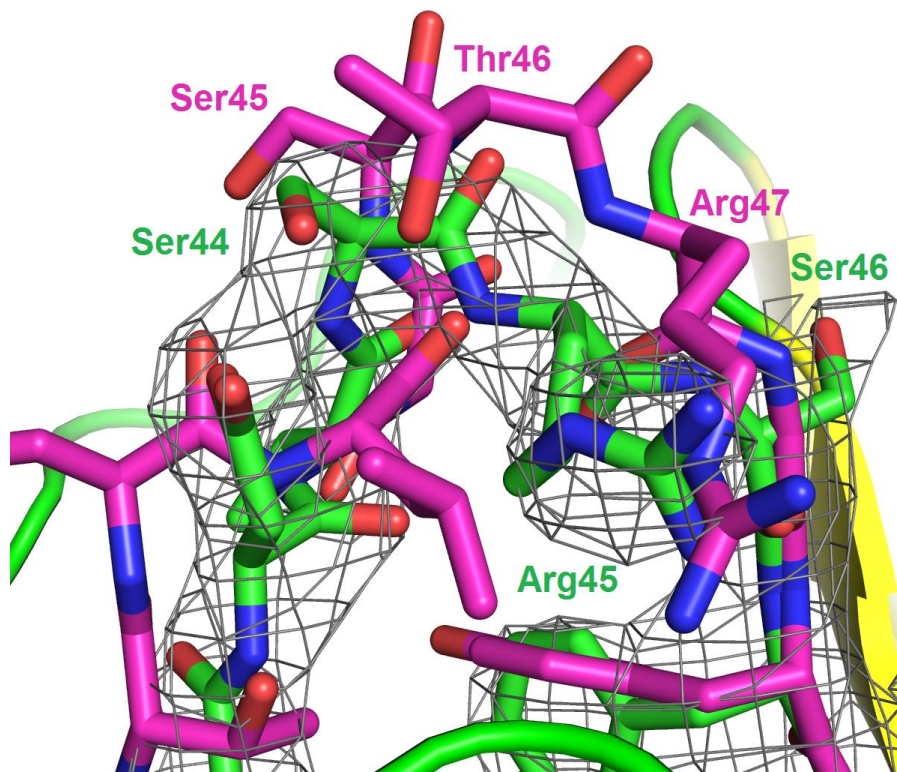
A resolution of less than 3.0Å is considered to be a low resolution data set (Headd *et al.*, 2012). Low resolution structure refinement and model building is akin to looking at an out of focus photograph, although the general scene may be deciphered, the fine detail is lost. At 3.2Å, it is frequently impossible to determine the exact orientation of peptide bonds, side chain rotamers and in many cases there is no clear electron density for the side chains, particularly for Lys, Arg and Glu residues. The solution to this problem is to build the protein backbone with idealised geometry based on prior knowledge of biochemistry. Poorly resolved side chain rotamers are placed in their most common conformation, but with increased B-factors or reduced occupancy to account for the lack of electron density. Such steps can make the results represent more a study of ideal geometry with limited crystallographic observations added. However, there is little option when presented with limited experimental data. The main chain of the protein can often still be built with reasonable confidence, particularly for a short pentraxin,

whose subunits are characterised by a well-defined tertiary structure and short flexible loop regions.

When refining low resolution data sets extra steps must be taken to prevent over-refinement. As mentioned in Chapter 2, the data-to-parameter ratio in crystallography is poor, even for high resolution data sets ( $<1.5\text{\AA}$ ). At  $3.2\text{\AA}$  resolution this issue is even more acute, so special tools and parameter modifications must be utilised. Reciprocal space refinement of rCRP was performed using Phenix applying group B-factor refinement with NCS and TLS restraints. TLS (translation, libration, screw) restraints treat whole domains or regions of a protein structure as groups whose motion is correlated by the fact that they are bonded to one another, therefore their motion is restricted within a range of possible translations, libration and screw axes (Winn *et al.*, 2001; Rupp, 2010). This can greatly reduce the number of parameters needed to refine a protein structure, making it particularly effective for low resolution refinement. Using group B-factors is another way of reducing the number of parameters needed to refine a structure with limited x-ray data. Group B-factors treat neighbouring atoms, such as side chain atoms in Lys, as likely to exhibit similar freedom of movement, instead of each atom's B-factor being adjusted individually, they are adjusted by the same value regardless of the differences in their individual starting temperature factor values.

The sequence of the protein model output from the molecular replacement was converted from human to rat in Coot using the mutate function to manually edit the amino acids based on the sequence of rCRP determined by Rassouli *et al.* (1992). The conversion of the amino acid sequence also required manual rebuilding of regions where the model deviated from the path plotted by the electron density map. The clearest example of this was the short loop between residues 44-50 (Figure 3.5). hCRP contains two extra residues Ser and Thr not present in the sequence of rCRP (Table 3.1). Ser45 and Thr46 were deleted from each pentamer using Coot, the chains were then renumbered using PDSet and real space refinement of Ser44 and Arg45 performed to bring the two residues within bonding distance. Ten calcium ions were added to the structure in the cation binding pocket identified for CRP and metal coordination

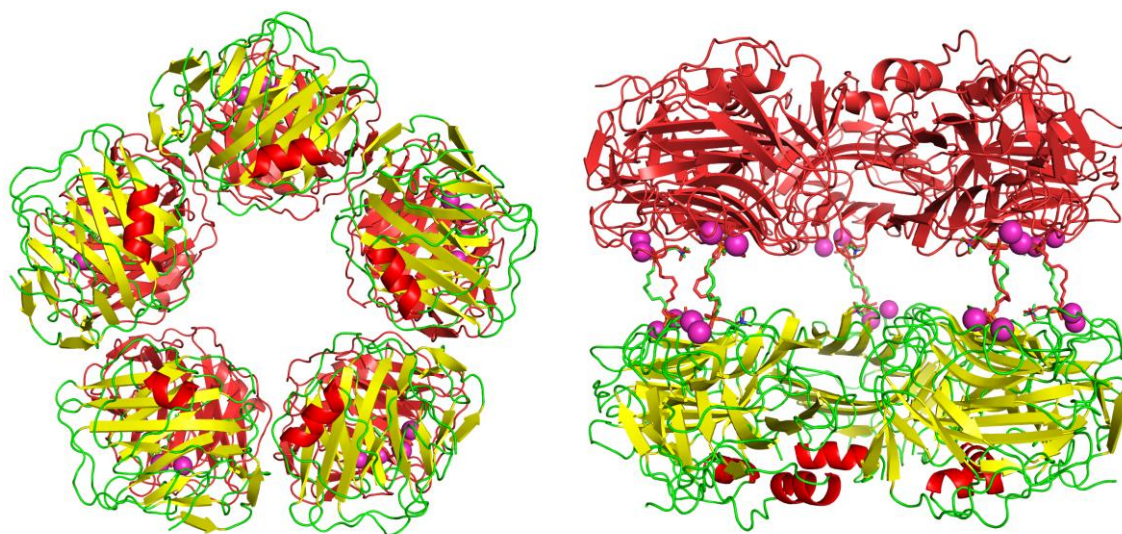
restraints were generated using phenix.readysset to assist positional refinement of the ions (Thompson *et al.*, 1999).



**Figure 3.5.** Stick representation of residues 42-46 of rCRP and the corresponding residues of hCRP. A key sequence difference in rCRP is the shorter loop between residues 42-46, hCRP has an extra Ser and Thr residue that are absent in rCRP. The original path followed by the hCRP search model (PDB: file 1B09; magenta chain; atom colours: carbon (magenta), oxygen (red), nitrogen (blue)) deviates from the path of the electron density map (grey mesh;  $1.5\sigma$ ). The green ribbon shows the model tracing the electron density after manual rebuilding of the rCRP amino acid sequence in Coot (green chain; atom colours: carbon (green), oxygen (red), nitrogen (blue)). Image generated with Pymol.

Five BPC8 ligands were built into the structure, but as the asymmetric unit contained only one pentamer, the ligand was built on a symmetry axis and therefore the occupancy was set to 0.5 to account for the symmetry-related ligand generated in each subunit as a result of the spacegroup (Figure 3.6).





**Figure 3.6.** Cartoon depiction of the x-ray crystal structure of rCRP-BPC8 complex. Left: Top-down view, right: side-on view showing BPC8 non-covalently cross-linking two pentamers. The cartoon depiction shows a  $\beta$ -jellyroll topology composed of two anti-parallel  $\beta$ -sheets (yellow) linked by short loops (green), with a short  $\alpha$ -helix (red) on the A-face of each protomer. The asymmetric unit of the rCRP crystal contains only a single pentamer in spacegroup  $P6_122$ . This means that a single decameric complex is placed either side of a two-fold rotation axis, the symmetry-related copy (red pentamer) was generated using the symmetry mate function in Pymol. BPC8 interacts with both pentamers either side of the symmetry axis, building the ligand into both binding sites generates a symmetry-related copy, resulting in two ligands per binding site in the model. As only one ligand can occupy a binding site at once, the occupancy is set to 0.5 and the symmetry copies treated as alternative conformers. Calcium ions are shown in magenta.

Continuous rounds of real-space model building in Coot and refinement in Phenix gave a final R-free of 27.74% and an R-work of 22.95%. The rms angle value of  $1.02^\circ$  and rms bond length of  $0.01\text{\AA}$  are half the upper limit outlined in Chapter 2, providing confidence that the geometry of the model has not be excessively influenced by the low resolution electron density, but equally the model shows some deviation from ideal geometry suggesting that the crystallographic data has influenced the model. The final model was analysed using the Molprobity structure validation server (Chen *et al.*, 2010). Molprobity analyses inter-atomic steric clashes, rotamers, Ramachandran angles and searches for unlikely bond geometry. As well as outputting data for each parameter, Molprobity aggregates the results of each parameter into a Molprobity score, which can be compared against all other structures of similar resolution deposited in the PDB to

give an indication of the model's relative quality. The rCRP-BPC8 model refined with riding hydrogens added had a clashscore of 12.3, which is in the 95% percentile, while the Molprobity score of 2.0 is in the 100<sup>th</sup> percentile, in both cases a lower score indicates a better result. The results from Molprobity are listed in Table 3.3.

Space Group	P6 <sub>1</sub> 22
Unit Cell Dimensions	a = 111.50, b = 111.50, c = 424.74 Å $\gamma$ =120.00°
Resolution Range	96.54 – 3.20 Å
No. of Reflections	924 260
No. of Unique Reflections	26 789
Multiplicity	34.5 (35.9)
Completeness (%)	99.7 (99.6)
R <sub>pim</sub>	0.05 (0.29)
(I)/ $\sigma$ (I)	14.3 (2.7)
Wilson B-factor	51.8
<b>Refinement</b>	
Molecular Replacement Model	1B09
No. of Residues	1020
No. of Solvent Molecules	0
No. of Ligand Molecules	5 (BPC8), 10 (Ca), 5 (NAG)
Rmsd Bonds (Å)	0.01
Rmsd Angles (°)	1.02
Residues in Favoured Regions (%)	93.57
Residues in Disallowed Regions (%)	0.20
<b>Average B-factor</b>	
Protein	70.90
Ligand	93.72
Solvent	N/A
R-factor (%)	22.95
R-free (%)	27.74

**Table 3.2.** Processing and refinement statistics for rCRP-BPC8 crystal x-ray diffraction data. Figures in parentheses denote results for the highest resolution (3.37-3.20Å).

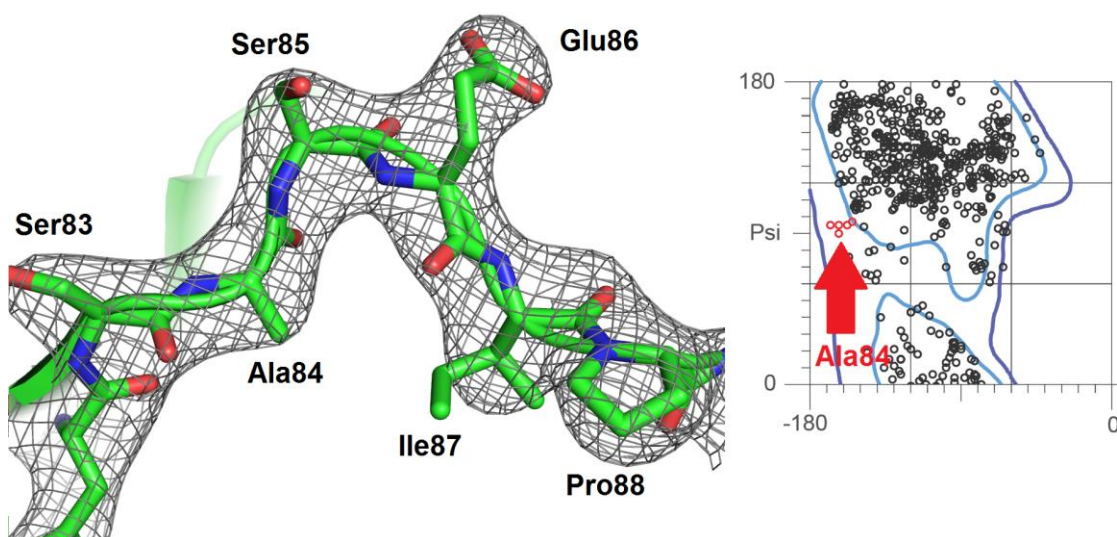
Clashscore	12.3
Poor Rotamers (%)	0.82 (7 residues)
Ramachandran Outliers (%)	0.20 (2 residues)
Ramachandran Favoured (%)	93.57 (946 residues)
Molprobability Score	2.02
C $\beta$ -deviations	0
Bad backbone bonds (%)	0
Bad backbone angles (%)	0

**Table 3.3.** Results of Molprobability analysis of final rCRP-BPC8 model. The percentiles indicated for the clash and Molprobability scores are relative to all structures deposited in the PDB with resolution between 2.94-3.44Å (N=1543). The Molprobability score is a combination of the clash score, Ramachandran and rotamer results normalised for the structures resolution, a detailed explanation of the criteria can be found in Chen *et al.*, 2010.

The Molprobability result, which causes the greatest concern is the low percentage of Ramachandran angles in the favoured region of the plot. Typically >98% is the targeted percentage of residues in the favoured region, but for the rCRP-BPC8 structure it is only 94.57%. Furthermore, 0.2% of residues are outliers in the forbidden region of the plot. The lower Ramachandran favoured percentage is likely a by-product of the low resolution and the refinement program's attempts to closely match the poorly resolved electron density. Residues in loop regions are particularly prevalent in the allowed area of the Ramachandran plots, for example Ala84 from the loop region formed by residues 83-90 is found in the allowed region for all subunits. This is expected as pentraxins are characterised by tightly associated secondary structure elements, which can impose strained geometry on the short loop regions linking the  $\beta$ -strands. The electron density map for residues 83-88 is shown in Figure 3.7 accompanied by the Ramachandran plot for the general case of residues in the rCRP-BPC8 structure, the position of each Ala84 is highlighted occupying the allowed region of the plot for all subunits. The low resolution means the electron density is somewhat amorphous, looking more like a tube in which the protein backbone is contained (Figure 3.5), rather than the detailed contour map seen for NADPro-SAP in Chapter 2 (Figure 2.13). The poor definition in the electron density map makes it difficult to build the protein backbone with certainty. The refinement programs will refine the position of the backbone to give an acceptable fit



with the best geometry. Ramachandran angles are not used in refinement, so can act as a less biased judge of final model. It seems that efforts to accommodate the model within the amorphous electron density has led to compromises on ideal geometry, giving more regions that are “allowed”, but not necessarily “favoured” as Ala84 in Figure 3.7 appears to show. It may be that these residues have strained backbone torsion angles by virtue of their position within the protein chain, but it cannot be known for certain from the rCRP-BPC8 data set due to the lack of definition in the electron density map. This limits the certainty of conclusions and means the electron density map and model built within should be viewed as a guide to the structure and side chains conformations viewed as the most likely result given the data.

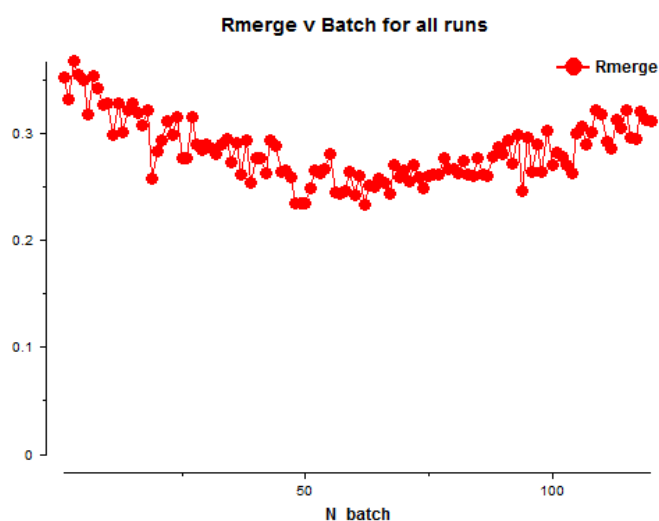


**Figure 3.7.** Left: 2Fo-Fc electron density map (grey mesh; contoured at  $1.5\sigma$ ) of residues 83-88 of chain B in rCRP-BPC8 generated in Pymol. Right: part of the Ramachandran plot for the general case of residues in the rCRP-BPC8 structure, the red arrow indicates the position of Ala84 from each subunit. The residues are seen within the allowed region (dark blue contour), but outside the favoured (light blue) portion of the plot. The low resolution of the data makes it difficult to determine the correct orientation of peptide bonds in the protein backbone, which may explain the lower % of backbone torsion angles falling into the favoured region of the Ramachandran plot as determined by Molprobity (Chen *et al.*, 2010). Model colours: carbon (green), oxygen (red), nitrogen (blue), green ribbon is cartoon representation of protein structure.

### 3.4. Discussion

#### 3.4.1. Radiation damage

The rCRP-BPC8 crystals still produced visible diffraction after a full dataset had been collected, overcoming the issue that plagued previous diffraction experiments. Special attention was paid to the extent of radiation damage during downstream processing of the diffraction data. If radiation damage was still an issue then it would be expected that the quality of diffraction would be impaired in the later diffraction images. Comparison of merging R-values for symmetry-related reflections on different diffraction images can be used to determine if crystals have been severely degraded by radiation damage. Radiation damage will diminish the intensity of previously recorded diffraction spots. A graph of merging R-values by image for a crystal affected by severe radiation damage will show an asymmetric, positive parabolic curve with values increasingly rapidly for later images. A graph of merging R-values against the image (batch) was plotted in Scala for the first data collection pass of the rCRP-BPC8 crystal (Figure 3.8).

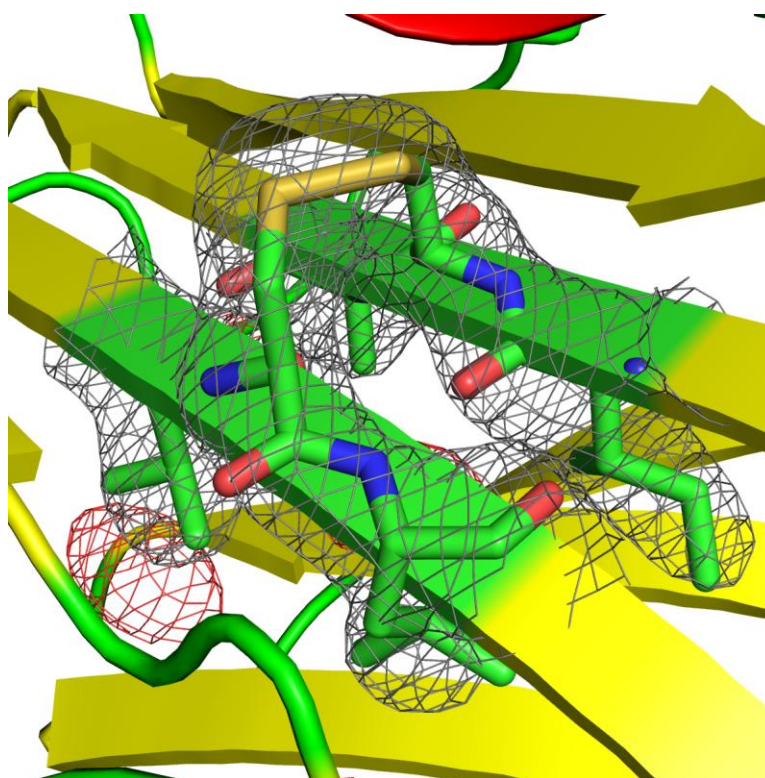


**Figure 3.8.** Graph of merging R-values for each image (batch) for rCRP-BPC8. Comparison of differences in intensity of symmetry-related diffraction spots can be used to assess the extent of radiation damage. Radiation damage does not appear to have significantly impaired the quality of recorded data as the R-merge at the end of data collection does not exceed that of the first few images. Image generated by Scala.

The graph of merging R-value against batch shows that the crystal has not suffered significant radiation damage. Larger R-merge values are seen for the first batches

because the population of data points is smaller, so outlying values are more influential on the average intensity. After reaching a minimum value, the R-merge increases as a result of progressive radiation damage. If the radiation damage is significant then the values will ramp up significantly and greatly exceed the values for the initial batches. In the case of the rCRP-BPC8 data, the later values were less than those recorded for the initial diffraction images, therefore radiation damage was not a limiting factor during data collection.

The electron density maps were also investigated for signs of radiation damage. Scission of disulphide bonds and side chain degradation of glutamate and aspartate residues are commonly observed as a result of local radiation damage (Weik *et al.*, 2000; Burmeister, 2000). The limited resolution of the data means that many of the side chains of Glu and Asp are poorly resolved, so they are not reliable indicators of radiation damage. However, the electron density of the intramolecular disulphide bond formed between Cys36-Cys95 on each subunit were analysed for evidence of radiation damage. If radiation damage has caused scission of the disulphide bond then a negative electron density peak would be observed in the Fo-Fc difference map indicating that the electron density predicted by the model exceed the electron density observed in the data. Figure 3.9 shows the typical electron density map observed for the disulphide bond.

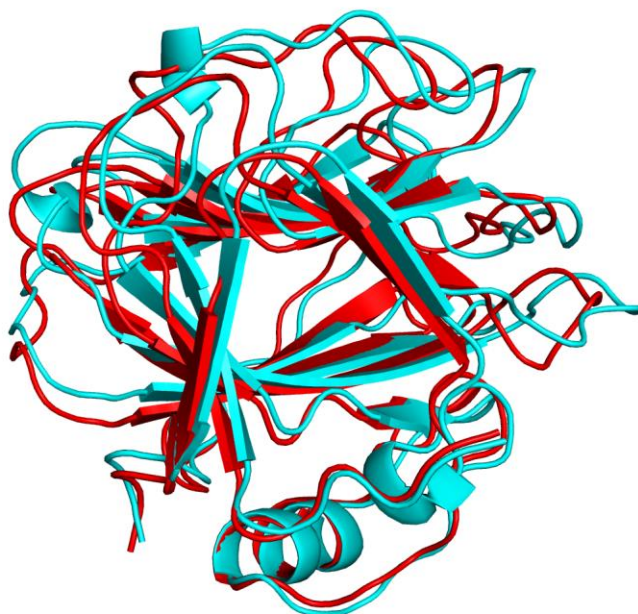


**Figure 3.9.** Electron density map of Cys36-Cys95 disulphide bond. 2mFo-DFc electron density map is represented by the grey mesh, Fo-Fc electron density “difference” map (red mesh indicating areas of the model where the expected electron density of the model differs from the observed data. Model atom colours: carbon (green), oxygen (red), nitrogen (blue), sulphur (yellow). There is no difference density around the disulphide bonds indicating that no significant bond scission was detected, suggesting radiation damage was not a significant issue for the rCRP-BPC8 x-ray diffraction data. There is a small region of negative difference density underneath the sheet, this was most likely caused by phase error in the structure factor calculations as there are no atoms are built into this region of the map.

Red electron density in Figure 3.9 indicates an area where the electron density of the model differs from the electron density observed in the diffraction experiment. No red electron density is observed in the area occupied by the disulphide bond, indicating that they have not been significantly disrupted by exposure to ionising radiation. The importance of this observation should not be underplayed, as some disulphide bond scission is commonly detected in data collected for the pentraxins. The absence of any detectable disulphide bond scission shows that the rCRP-BPC8 crystals are relatively resistant to radiation damage, which is in contrast to previous observations of rCRP crystals.

### 3.4.2. Structure analysis

The overall structure of rCRP remains consistent with the observations for the hCRP (Thompson *et al.*, 1999). The protein consists of five identical subunits arranged in a pentagonal radial symmetry (Figure 3.6). The subunits have two anti-parallel  $\beta$ -sheets folded into a flattened  $\beta$ -jellyroll topology. A short  $\alpha$ -helix is present on the so-called 'A-face' of the subunit, while on the opposite 'B-face', there is a double calcium binding site in a shallow gutter formed by the curvature of one  $\beta$ -sheet. The protomers of rCRP show a slight rotation compared to those of hCRP with calcium binding site tipped approximately  $2^\circ$  towards the central pore of the pentamer and a centre of rotation at the  $\alpha$ -helix (Figure 3.10).

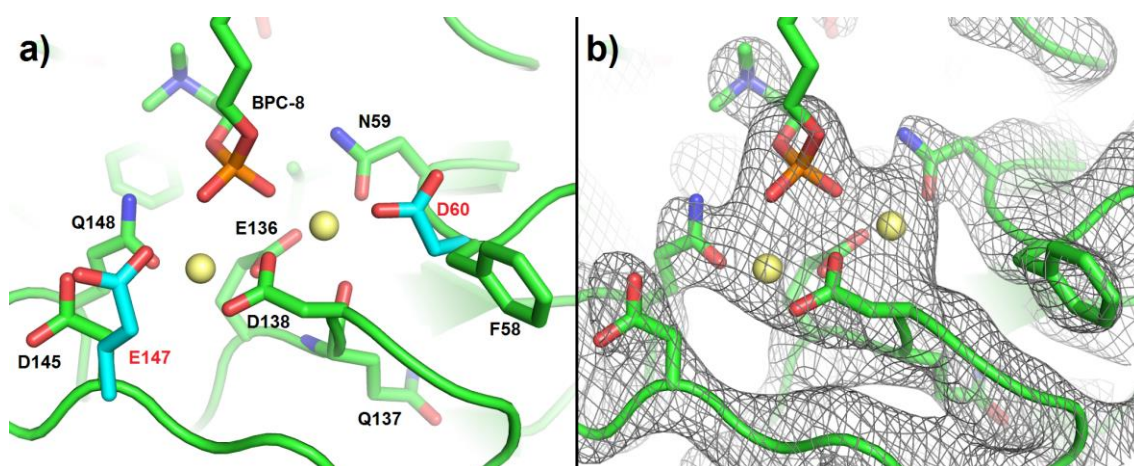


**Figure 3.10.** Single protomer of rCRP (red) and hCRP (turquoise; PDB file: 1B09) viewed along the axis of the  $\beta$ -jellyroll. The protomer of rCRP is rotated by around  $2^\circ$  moving the calcium binding site at the top of the figure towards the central pore of the pentamer, with the centre of rotation at the  $\alpha$ -helix at the bottom of the image.

Two calcium ions, separated by  $4\text{\AA}$  are coordinated by an amino acid side chains from nearby loops (Figure 3.11). The phosphocholine ligands interact with the protein by two oxygens of the phosphate coordinating the calcium ions. The calcium binding



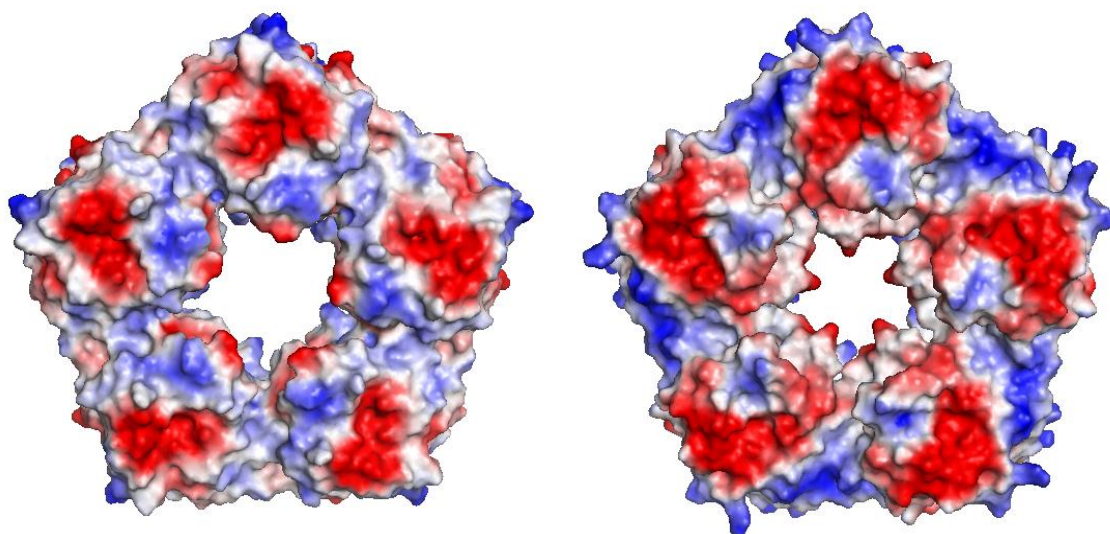
environment is slightly different to that of human CRP. The coordination of the first calcium ion is reduced by the substitution of Asp58 in hCRP by Phe58 in rCRP. Similarly, coordination of the second calcium ion is reduced from four amino acid side chains in hCRP to three in rCRP by replacement of Glu147 with Asp145. These differences are highlighted when Asp58 and Glu147 from hCRP (PDB file: 3L2Y; Mikolajek *et al.*, 2011) are superimposed on the calcium binding site shown for rCRP in Figure 3.11a. Figure 3.11b shows clearly that the electron density for Asp145 points away from the metal binding pocket in the same way observed for SAP. Based on the fact that larger metal ions have been observed binding in the second calcium binding site for hSAP (Emsley *et al.*, 1994), it is conceivable that because the coordination of both calcium ions are reduced in rCRP that both sites may be capable of binding larger metal ions.



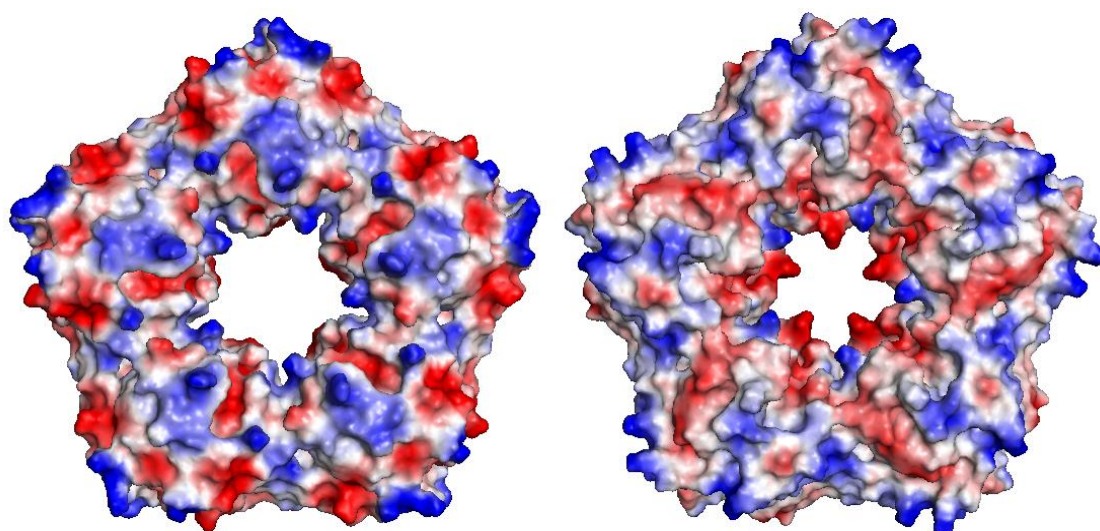
**Figure 3.11.** (a) The calcium binding site of rCRP in complex with BPC8 (green chain). Turquoise residues, Glu147 and Asp60, have been superimposed from hCRP (PDB: 3L2Y) (turquoise chain) to highlight the reduced coordination of both calcium ions (yellow) in rCRP compared with its human homologue. (b) 2Fo-Fc (Adams *et al.*, 2010) electron density contoured at  $1.5\sigma$  (grey) fitted with the double calcium binding site and surrounding residues.

### 3.4.3. Surface charge

Diagrams of the surface charge of the A- and B-faces of rCRP and hCRP were generated using Pymol (Pymol Molecular Graphics System, Schrödinger LLC). The extent and nature of inter-protein contacts will influence packing and stability of crystals produced. Observing the differences between the surface charge distribution of rCRP and related pentraxins may highlight the cause of the radiation sensitivity observed in previous studies. Figure 3.12 depicts the surface charge on the B-faces of rCRP and hCRP, with Figure 3.13 depicting the charge on the A-faces. rCRP could be compared to other pentraxins as hCRP is not a simple protein to crystallise, crystals of hSAP that diffract to high resolution ( $<1.5\text{\AA}$ ) have been obtained. The highest resolution structure of hCRP deposited in the protein data bank is  $2.5\text{\AA}$  (PDB code: 1B09; [www.rcsb.org/pdb](http://www.rcsb.org/pdb)). However, hCRP crystals do not typically suffer from the hyper radiation sensitivity observed for rCRP, so it seems an appropriate choice for comparison as it is most closely related human protein.



**Figure 3.12.** Surface charge map of the B-faces of left: rCRP and right: hCRP. Red areas of the map indicate acidic residues, with blue the basic residues and white signifying non-polar residues.



**Figure 3.13.** Surface charge map of A-faces of left: rCRP and right: hCRP. Red areas of the map indicate acidic residues, with blue the basic residues and white signifying non-polar residues.

There is a clear reduction in both positive and negative charges on the B-face of rCRP compared to its human homologue. This is clearest around the calcium-binding site of the protein, where fewer acidic residues at the surface make the site appear smaller. However, it is also clear that there are fewer basic surface residues and that the strength of the positive charge is less for rCRP than hCRP. This is most apparent towards the outer edge of the pentamer (Figure 3.12). In hCRP, a clear region of positively charged residues between the binding sites of neighbouring protomers make up a significant portion of the outer edge of the pentamer. In rCRP, the same area of the pentamer is a mix of mildly basic, mildly acidic and non-polar residues that cannot be clearly defined as an either positively or negatively charged region.

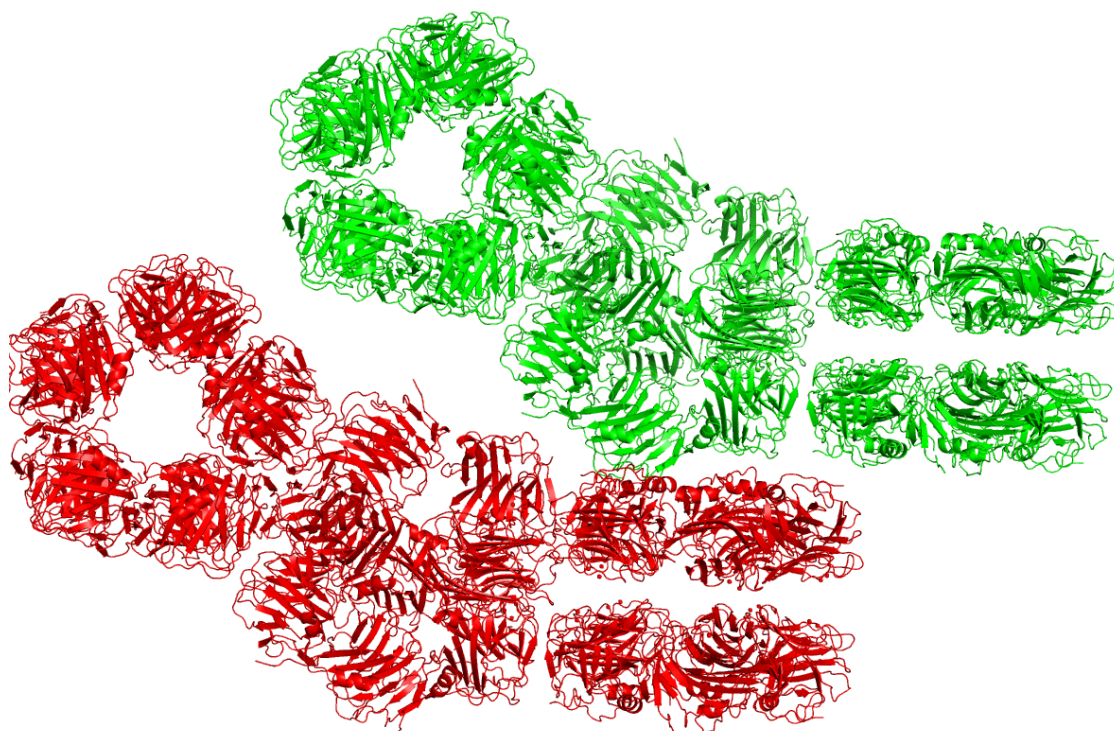
The A-faces of both proteins depicted in Figure 3.13 shows a more complex patchwork of small areas of alternating positive and negative charge. It is difficult to make clear inferences from the maps, but it appears that rCRP shows more areas of strongly acidic residues than hCRP. This is particularly pronounced at the outer edge of the pentamer where two small areas of negative charge are discernible either side of the small positively charged region at each vertex of the pentamer. Analysis of the crystal packing shows how the surface charge distribution has influenced inter-protein contacts.



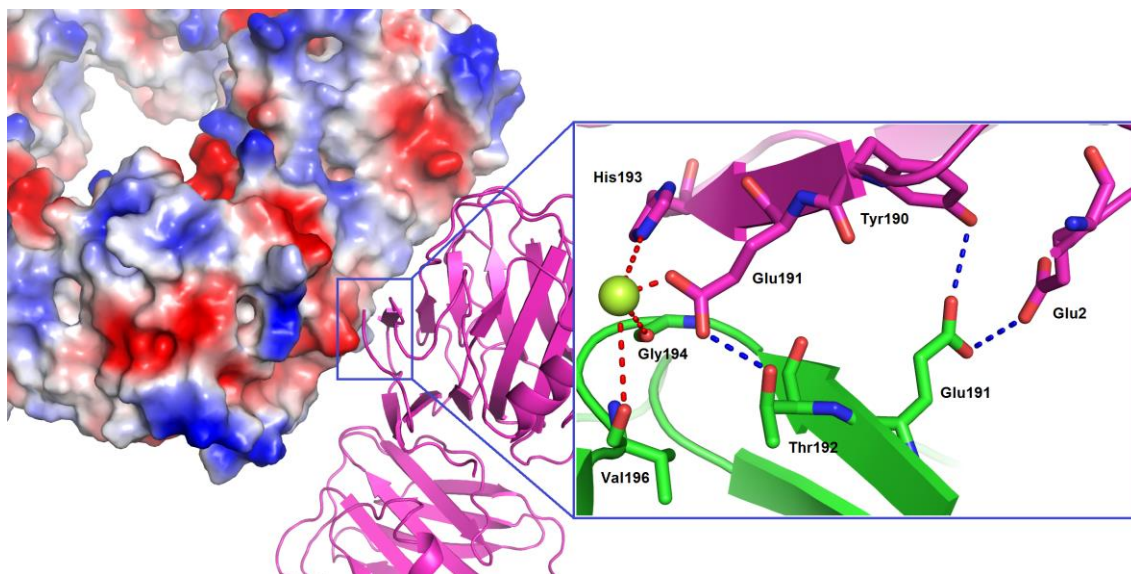
#### 3.4.4. Crystal packing

The packing of rCRP in the crystal is greatly influenced by the presence of the bivalent ligand. BPC8 induces decamer formation through non-covalent cross-linking rCRP in a B-face to B-face configuration (Figure 3.6). There are no significant interactions between surface residues on the opposing B-faces of the decamers. It is difficult to conclude whether this is because it is energetically unfavourable or because the length of the central alkyl chain of the ligand prevents closer contact between pentamers in the decameric complex. Although, it would be expected that if there were a number of favourable B-face to B-face interactions then the flexibility of the alkyl chain would allow it to fold, bringing the pentamers into closer contact. Therefore, it seems reasonable that no favourable contacts exist between opposing rCRP B-faces when decamerised by BPC8.

The use of BPC8 means that any inter-protein contacts between neighbouring decamer complexes are limited to the A-faces and outer edges of the protomers. Using molecular graphics program, Pymol, it is possible to generate the crystal packing of rCRP-BPC8 pentamers (Figure 3.14). The decamers appear as if arranged in long twisting non-covalent chains, with each decamer rotated by 45° from the previous pentamer. This requires multiple inter-decamer contacts at the outer edge of the complex. Four protomers from each decamer are involved in inter-protein interaction with a neighbouring complex. The rotation of one complex relative to its neighbour means that two of the four protomers form two areas of interactions, with the other two forms only a single inter-protein interaction. The fact that the complex is a pentagonal prism means the chains of decamers pack into a twisted arrangement of decamers.

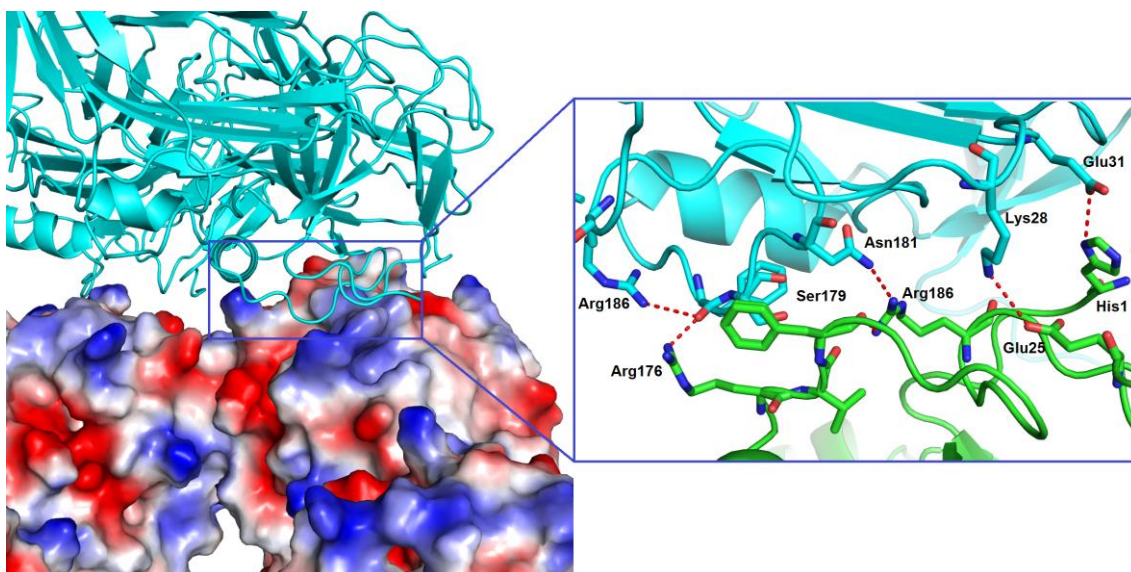


**Figure 3.14.** Crystal packing of rCRP decamers. Decamers appear as if formed into long twisted non-covalent chains by interactions at the edges of protomers in neighbouring complexes. The rCRP pentamers are coloured depending on the chain to which they belong, BPC8 has been omitted for clarity. There is little interaction between red and green chains, but some interaction between the A-faces of two pentamers from different chains can be seen.



**Figure 3.15.** Example of the potential interprotein contacts formed between neighbouring decameric complexes. Left: surface charge map of a single rCRP pentamer generated in Pymol with close contacts to a neighbouring pentamer (magenta cartoon) highlighted by the blue box. Right: enhanced image of close inter-protein contacts between one pentamer (green ribbon, green carbons, red oxygens and blue nitrogens) and the neighbouring pentamer (magenta cartoon, magenta carbons, red oxygens, blue nitrogens). Red dashed lines show potential inter-protein contact bridged by a coordinated calcium ion (limon yellow; calcium ion may be present but electron density not of sufficient resolution to be built into model with certainty), blue dashed lines show potential hydrogen bonding.

There seems to be little interaction between different decamer chains pictured in Figure 3.14. Figure 3.14 depicts two separate arrangements of rCRP-BPC8 decamers. There appears to be only one point at which the red and green chains come into close enough contact for an inter-protein interaction. This interaction involves the A-faces of two pentamers, with close association of residues from two protomers of each pentamer (Figure 3.16). Generating a surface charge map shows that this interaction involves the mostly non-polar region (residues of one pentamer) interacting with the mostly basic region in the middle of the A-face of rCRP (Figure 3.13). It is expected that this interaction is mediated by solvent or small molecules from the crystallisation cocktail, which cannot be discerned in the electron density map due to its limited resolution.



**Figure 3.16.** rCRP A-face charge map interaction. Left: surface charge map of the A-face of one rCRP pentamer generated in Pymol, blue box shows area of close contact with the A-face of a pentamer from a separate decamer strand (turquoise cartoon). Right: enhanced image of potential inter-protein contacts (red dashed lines) between residues on the loop region of one pentamer (green cartoon, green carbons, red oxygens and blue nitrogens) and a loop region from another pentamer (turquoise cartoon, turquoise carbons, red oxygens and blue nitrogens).

It is clear that the observed crystal packing would not be possible without the use of BPC8 because of the lack of interaction between the B-faces of rCRP pentamers involved in the complex. It would not be possible to form the rCRP decamers that are the basis of the crystalline structure. The A-face to A-face interactions would be possible in a crystal not containing decamers of rCRP, but it seems highly unlikely that this alone would stabilise crystal formation. From the successful data collection, it is possible to conclude that the use of cross-linking agent has produced crystals of rCRP with unique, stable inter-protein contacts. The ligand itself mediates stable inter-protein interactions and facilitates others by providing otherwise unobtainable decameric architecture.

### 3.4.5. Extended C-terminus

Both rat pentraxins have a short extended C-terminal region compared with pentraxins from other species (Rassouli *et al.*, 1992; Dowton and McGrew, 1990). rSAP is extended by the sequence NKSS, similar to the NKIS sequence present on Armenian Hamster SAP (Dowton and Waggoner, 1989). In contrast, the seven-residue C-terminal fragment of rCRP, LTDCCES, is highly unusual amongst mammalian pentraxins. The extended C-terminus of rCRP is reported to form three intra-molecular disulphide bonds per pentamer with the termini of the other two subunits forming intermolecular disulphide bonds. Non-reducing SDS PAGE gives two bands one around 25kDa corresponding to a single subunit and a 50kDa band corresponding to two subunits (Rassouli *et al.*, 1992). The electron density at the C-terminus of the rCRP crystallographic data presented here is sparse and uninterpretable suggesting the region is highly flexible. Furthermore, no evidence was observed for an intermolecular disulphide bond, but this could be because the pentameric symmetry of the protein means there are five possible orientations for the single intermolecular disulphide so averaging over all orientations means none are visible at the limited resolution of the data. For this reason neither an extended C-terminus nor an intermolecular disulphide bond have been built in the rCRP model presented here.

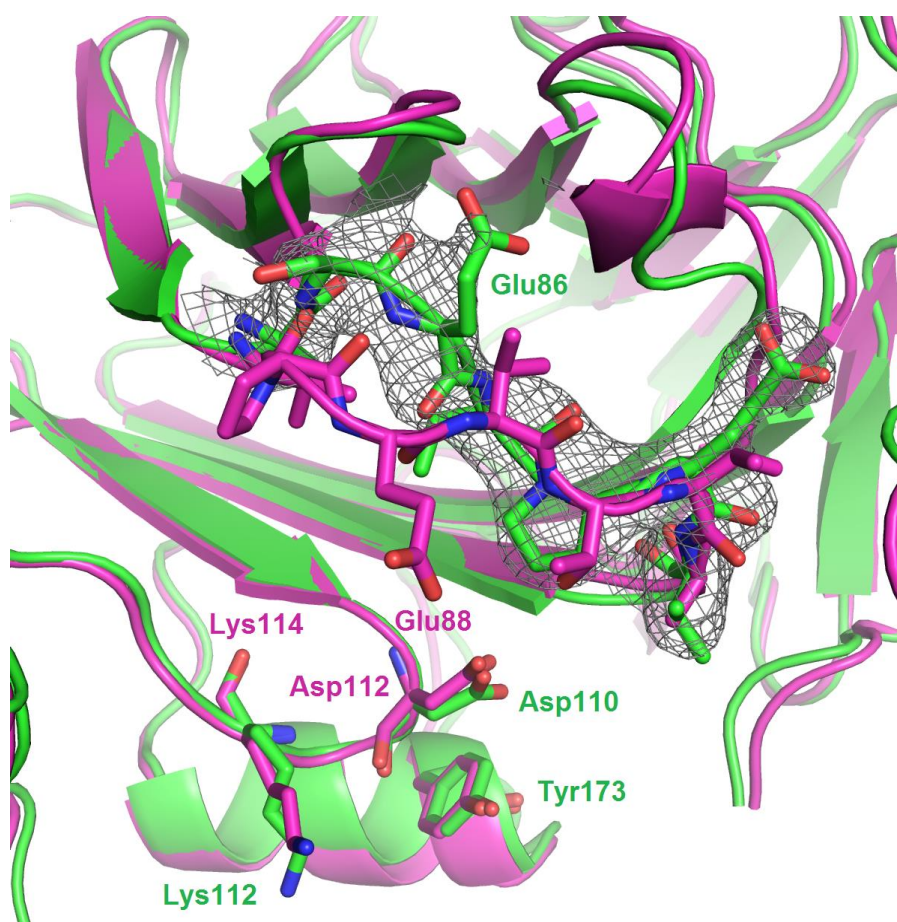
### 3.4.6. Complement activation

hCRP activates the complement system by the classical pathway by binding C1q of the C1 complex through clefts on the A-face of each CRP protomer (Kaplan and Volanakis, 1974; Agrawal and Volanakis, 1994). Complement activation occurs when the B-face of CRP binds to one of its ligands, such as chromatin, histones and nucleosome core particles. This strongly implies a role for CRP in coordinating the removal of nuclear material from dead or dying cells by phagocytes. Similarly, CRP is known to activate complement by binding to the C-polysaccharide of *streptococcus pneumoniae* potentially demonstrating its role in innate immunity (Abernathy and Avery, 1941). Oddly, it has been reported that rCRP does not activate complement when binding to C-polysaccharide of *streptococcus pneumoniae* (de Beer *et al.*, 1982). However, binding of rCRP to other ligands such as phosphocholine has been shown to activate complement (Padilla *et al.*, 2003). These findings suggest that differences in the

structure of rCRP from hCRP prevent or reduce complement activation when binding to particular ligands. It could be that the ability to bind to foreign pathogens is lost, but the ability to bind nuclear material is retained.

The cleft on the A-face for C1q binding contains a number of residues identified by mutagenesis as being important for complement activation by hCRP (Agrawal and Volanakis, 1994; Gaboriaud *et al.*, 2003); in particular, residues Glu88, Asp112, Lys114 and Tyr175. The analogous residues in rCRP are Glu86, Asp110, Lys112 and Tyr173. The rCRP-BPC8 structure was investigated to find any clear deviations with a structure of hCRP downloaded from the protein data bank (PDB code: 3KQR). Figure 3.17 shows the cleft on the A-face of rCRP with the residues of interest highlighted; superimposed are the key residues of hCRP with some surrounding residues included for clarity. It is apparent that the only noticeable difference in the position of these residues is between Glu86 of rCRP and Glu88 of hCRP. In hCRP, the loop of which Glu88 is a part bends towards the central pore of the pentamer. In contrast, the loop of rCRP stays closer to the surface of the protomer. This best displayed by the distance of the C $\alpha$  of Trp65, which is part of a  $\beta$ -sheet, and Glu86, both of which are conserved in hCRP as Trp67 and Glu88. In hCRP, the distance between the C $\alpha$  is around 13Å, while in rCRP the distance is closer to 10Å. The shift in position appears to be caused by Pro88 on rCRP which causes the loop to fold towards the protomer. However, it must be considered that in both proteins this region is a loop and likely to show greater than average flexibility. The loops may be able to fold in different conformations, but x-ray diffraction recorded only a single conformation because the low temperature at which the data is collected reduces the thermal motion of the loop. It is therefore possible to suggest that this difference could affect the affinity of binding to C1q as it is the only key complement binding residue showing deviation in position from its homologous residue in hCRP. However, the suggestion is tempered by the limited resolution of the data and the position of Glu86 on a potentially flexible loop region of the protein.





**Figure 3.17.** Complement binding cleft of hCRP (magenta; PDB file: 1B09) and rCRP (green) with residues identified as important for complement binding, as identified by Gaboriaud *et al.*, 2003, annotated. The grey mesh 2Fo-Fc electron density map shows how the position of the Glu86 loop in rCRP differs from the corresponding position determined for the crystal structure of hCRP in PDB file 1B09.

Another structural difference that may influence the affinity of complement binding is the extended C-terminal LTDCCES region. As discussed in the previous section, there was no clear electron density in which to build the extended C-terminus of rCRP. It was concluded that this indicated high flexibility in the six residues. The final residue that could be built into the structure was Pro204, whose C $\alpha$  is just 10Å from the C $\alpha$  of Tyr173. It must therefore be considered that this highly flexible portion of the protein could obstruct binding in the complement binding cleft. Based on a backbone length of 3.8Å per residue, the extended C-terminus could have an effective length of 22.8Å. In the case of LTDCCES, there is reportedly an intramolecular disulphide bond between the two cysteine residues causing a fold in the chain that would reduce the effective length, perhaps to between 11.4 and 15.2Å, which would still be long enough to reach

the complement binding cleft. Cleavage of the extended C-terminus or removal by mutagenesis would show definitively whether the extended C-terminus does influence complement binding, but the crystal structure shows that it is in close enough proximity.

There is significant variation between homologous pentraxins of different species in terms of behaviour as an acute phase protein, plasma concentration, glycobiology, ligand binding specificity and secondary ligand binding effects such as autoaggregation and complement activation (Baltz *et al.*, 1982; Pepys and Baltz, 1983). X-ray crystallographic analysis of the rat pentraxins reveals how the subtle structural differences between pentraxins from different species can result in significant shifts in ligand binding specificity.

The structure of rCRP described constitutes the first reported structure of the pentraxin from the rat. Previous observations of the high radiation sensitivity of crystals have been overcome. Further investigation is required to determine whether the use of bivalent ligands to non-covalently cross-link protein complexes has potential to solve similar crystallisation problems for other proteins.

The structure of rCRP shows that the coordination of both calcium ions is reduced relative to its human homologue. Similar observations for SAP have been shown to allow larger metal ions to fit into the more loosely bound metal site (Emsley *et al.*, 1994). It is therefore possible to speculate that both metal binding sites rCRP may be capable of incorporating cations larger than calcium, although this would need to be confirmed experimentally. It might be suggested that for a calcium binding protein that circulates with a normal plasma concentration of around  $450\mu\text{g ml}^{-1}$  the ability to incorporate a variety of metal ions would be of significant benefit to an organism such as the rat which has a highly varied diet and the constant bioavailability of calcium cannot be relied upon.

Some of the other previously reported unusual structural features are not well resolved by the x-ray diffraction data; in particular, the glycosylation and the extended C-terminus. Small amorphous different density is visible near glycosylation site Gln128 on



each subunit on the pentamer, but it is not possible to definitively place even a single saccharide unit. A similar situation is often seen in hSAP at lower resolution so this is not considered particularly unusual; a single saccharide unit at each glycosylation site has been built in the model for illustrative purposes, but with occupancy set to 0 in the PDB file. No evidence is present in the electron density maps for the extended C-terminal amino acid sequence LTDCCES or the reported intermolecular disulphide bridge between two of the subunits' extended C-termini. It could be that this region is highly flexible, something supported by the similar situation seen for the NKSS segment at the C-terminus of rSAP, where the B-factors are substantially higher than elsewhere in the structure (Mikolajek thesis, 2008). Proline-204 is the final residue before the extended C-terminal section of rCRP. Pro-204 is 24Å from the nearest Pro-204 on neighbouring subunits, so it may be possible for two seven residue fragments on neighbouring subunits to come into contact with sufficient overlap for disulphide bond formation, but there would not be much slack in the chain. We might expect this to cause visible distortion in the bond angles of Pro-204 to allow this overlap, but this has not been observed. More effort is required to determine whether there is a natural intermolecular disulphide bridge present in rCRP or whether it forms as an artefact during sample preparation for non-reducing SDS PAGE.

#### **3.4.7. Multivalent ligands as tools for crystallography**

It is expected that in previous studies the crystals of rCRP obtained involved pentameric protein complexes, seemingly as a result the crystals were highly sensitive to radiation damage. By forming decameric complexes radiation sensitivity has been overcome and while it is not possible to say that this is exclusive due to the use of bivalent cross-linking agents, it is clear that their use has made different inter-protein contacts possible and provided strong non-covalent interactions that stabilise the decameric assembly.

Further consideration should be given to the use of multivalent compounds to overcome crystallisation problems in x-ray crystallography. Multivalent ligands have been a feature of x-ray diffraction experiments involving the pentraxins for the past decade, with decameric structures of both SAP and CRP now known (Pepys *et al.*, 2002; Ho *et al.*, 2005; Kolstoe *et al.*, 2014; CRP unpublished structure). These structures have

markedly different crystal packing and inter-protein contacts compared to structures in complex with univalent ligands. Comparison of the crystal packing between the crystal structure of SAP in complex with CPHPC (PDB file: 4AVT) and the pentameric SAP-N-acetyl D-proline structure presented in Chapter 2 confirms that different inter-protein contacts exist when a bivalent ligand is used to non-covalently cross-link two SAP pentamers. 4AVT is a more representative comparison of structures because the crystallisation conditions used were similar to those used to crystallise the SAP-NADPro structure, ruling out the possibility that different crystallisation conditions were responsible for the difference in inter-protein contacts. In NADPro-SAP, the crystal packing involves close contact between parts of the A-face and the B-face of SAP (Chapter 2, Figure 2.11). This is impossible in a decameric SAP complex because the B-faces of the protein are buried in the centre of the decamer.

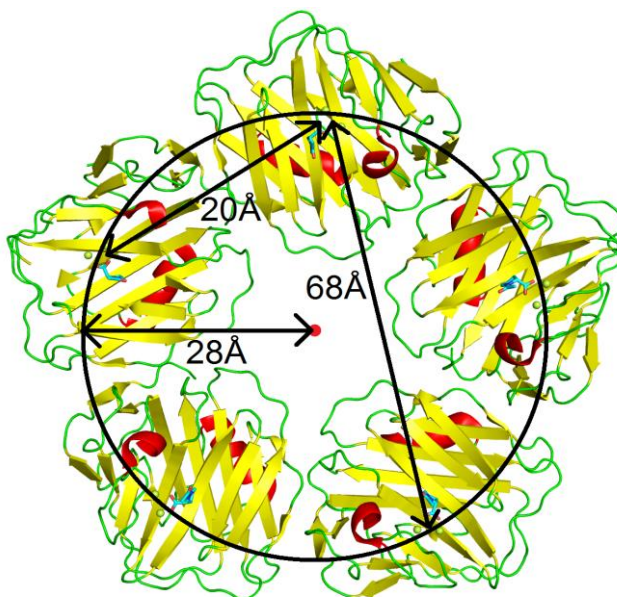
There is potential for the synthesis of multivalent molecules for a range of proteins whose structure remains undetermined or for which crystal structures are only available at limited resolution. Although, the resolution of the rCRP structure presented here is limited to 3.2Å, a decameric structure of SAP in complex with bivalent ligand CPHPC is deposited in the protein data bank with a resolution of 1.6Å (PDB code: 4AVV; [www.rcsb.org/pdb](http://www.rcsb.org/pdb)). This shows that forming a larger decameric protein-ligand complex does not hamper the achievable resolution, but in the case of rCRP it does allow for successful diffraction experiments.

## Chapter 4. Synthesis of multivalent ligands for SAP

### 4.1 Introduction

#### 4.1.1. The design of multivalent ligands for SAP

In addition to the design requirements determined in Chapter 2, there are a number of other key considerations for the synthesis of multivalent ligands for SAP. Firstly, the ligand needs to be water soluble for isothermal titration calorimetry and protein crystallisation studies. The designed molecule must be large enough to present headgroups to all five of SAP's binding sites simultaneously. Using the molecular graphics program, Coot (Emsley and Cowtan, 2004) it is possible to measure the distances between binding sites in SAP using crystallographic structure presented in Chapter 2. The double calcium binding site of each protomer is separated from the binding site on the neighbouring protomer by 20Å (Figure 4.1). The maximum distance between two binding sites is 68Å for the protomers on opposite sides of the pentamer. Drawing a circle whose circumference intersects all five binding sites it is possible to determine that a multivalent ligand would need a diameter of at least 56Å to target all five binding sites simultaneously. Finally, it is worth considering that many different ligands are known for SAP, so the multivalent scaffold and the synthetic strategy should be designed to allow ligation of different headgroups, if necessary.



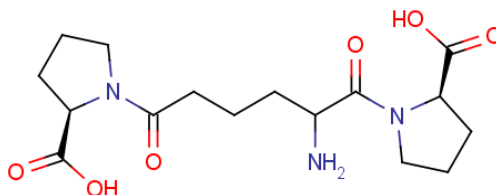
**Figure 4.1.** Cartoon representation of SAP pentamer demonstrating distances between ligand binding sites. Image drawn with Pymol (Schrödinger LLC.)

A multivalent ligand consists of three main parts. Firstly, the headgroup, which should be a univalent ligand that can easily be linked to other similar univalent ligands through a scaffold without compromising the mode of binding. Secondly, the scaffold, which is defined as the central component to which the headgroups are attached. Finally, the linker or arms used to link the headgroup to the central scaffold and provide sufficient separation between the headgroups and the central scaffold for simultaneous interactions with all binding sites of the target. The design choices made for each component will be briefly discussed.

#### 4.1.1.1. Headgroup design

Chapter 2 demonstrated that D-proline was a suitable headgroup for a multivalent ligand targeting SAP. Furthermore, evidence from Pepys *et al.* (2002) shows D-proline to have moderate univalent binding affinity, which is greatly improved by creating a bivalent analogue. This investigation seeks both pentavalent and decavalent ligands for SAP using univalent ligand, D-proline and bivalent ligand, CPHPC developed by Pepys *et al.* (2002). CPHPC contains no functional groups that could conceivably be used to attach to a multivalent structure without affecting the efficacy of binding, so a slight modification is needed. Replacing the hexanoyl chain of CPHPC with aminoadipic acid

(hGlu) introduces an amino group which can be used for ligation to a multivalent ligand (Figure 4.2). Although the modification does introduce a chiral centre into the molecule, the addition of an amino group was considered to be an appropriate modification that minimised structural differences between the new amino-CPHPC structure and the original bivalent compound upon which it was based.



**Figure 4.2.** Modified CPHPC headgroup amino-CPHPC, designed to allow ligation through the amino group while leaving the C-termini free to interact with SAP.

#### 4.1.1.2. Linker design

The design of the linker was influenced by a number of factors. The NADPro crystal structure presented in Chapter 2 showed some evidence that the acetyl group of NADPro favoured a *cis* conformation with a weak hydrogen bond between the carbonyl oxygen of the acetyl group and the amide side chain of Gln148. Therefore, the design of the linker for SAP should use an amide bond to ligate the headgroup to the rest of molecule and allow maximum flexibility for the headgroup to orientate and bind to SAP in its preferred conformation. As discussed in section 4.1.1.1, hGlu was chosen as an amino-containing analogue for the hexanoyl linker of CPHPC, this allows it to be attached to the linker through an amide bond. Glycine was chosen as the residue through which the headgroups would be attached to the linker because its structure is achiral and it lacks a side chain functional group, both of which may have interfered with binding of the headgroup to SAP.

The binding affinity of any multivalent ligand will be affected by the length of the linker; too long and the increased degrees of freedom will reduce the strength of binding; too short and a single ligand molecule will not be able to occupy all binding sites simultaneously. However, the ideal length for the linker is not immediately obvious because interaction between the linker and the surface residues of SAP will affect how the flexible chain folds. Therefore, it is best to produce a range of ligands that sample various lengths close to but in excess of the minimum required radius of 28Å. For this

reason polyethylene glycol (PEG) oligomers were chosen as a key constituent of the linkers. PEG-containing amino acids are commercially available in a variety of lengths; with a C-terminal carboxylic acid and Fmoc-protected N-terminal amino group they can be easily incorporated into a peptide-based structure. The added advantage of PEG is that it increases the water solubility of peptides, thus serving to fulfil another requirement of the multivalent ligands. However, experience with PEG-containing peptides has shown that they are often oily in nature, which makes them more difficult to handle. For this reason two gamma-aminobutyric acid (GABA) residues were introduced into the linker structure either side of the PEG group to balance the hydrophilicity of the compounds. The final design consideration was the functional group used to ligate the linker to the central scaffold, this could only be determined by first deciding upon the scaffold structure.

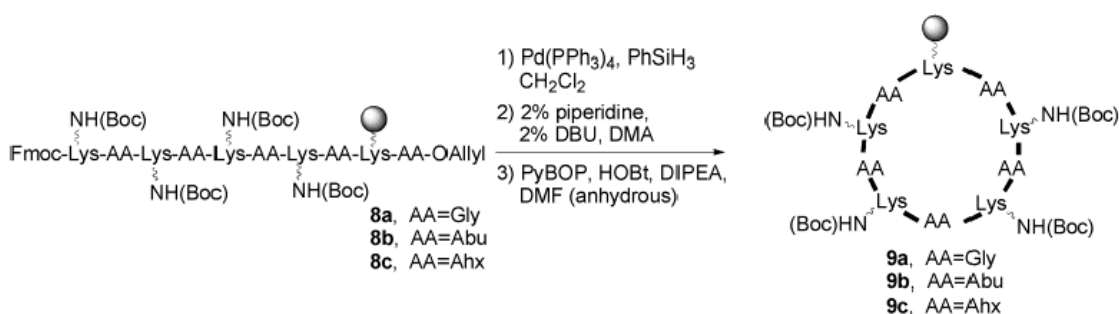
#### **4.1.1.3. Scaffold design**

Many different scaffolds have been employed in previous multivalent ligands studies. Kitov *et al.* (2000) utilised a monosaccharide as a scaffold in their synthesis of STARFISH multivalent binder of Shiga-like toxins and a plethora of examples of peptide and cyclodextrin-based multivalent ligands abound in the literature (reviewed by Niederhafner *et al.*, 2008; Deniaud *et al.*, 2011; Qu *et al.*, 2014). However, many of the cyclodextrin-based ligands are hexavalent and pentameric cyclodextrins are not widely commercially available. Scaffolds of dendrimers formed from polymers such as commercially available PEG and peptides like the multiple antigen peptide system developed by Tam (1988) were also considered. However, dendrimers were rejected for this study as it was thought a more rigid central scaffold would reduce the entropy penalty from binding, which should give an increased binding affinity. Research of the literature found that cyclic peptides with pentameric symmetry have been developed as the basis for multivalent ligands by Zhang *et al.* (2004). In their investigation, a central cyclic peptide scaffold was used, with multiple branches attached through the amino side chains of five lysines and a Galactose-based headgroup presented at the end of each branch. The group found the binding affinities of their peptide-based multivalent ligands to be as much as 100,000 fold stronger than univalent galactose. Zhang *et al.* also conducted computer simulations showing that the cyclic peptide maintained an

expanded cyclic shape and did not collapse into a folded secondary structure. For these reasons, the lysine-based cyclic peptide developed by Zhang *et al.* was determined to be an appropriate starting point for the design of multivalent ligands for SAP as it was likely to offer a rigid core of adequate size to counterbalance the flexibility of the PEG-based linkers.

However, the method of ligating the linkers to the central scaffold was modified from that of Zhang *et al.* Their method forms attachments between the cyclobutenyl groups on the linker and the lysine side chains of the scaffold. As the cyclobutenyl group was not readily available, the side chain lysine functional groups were modified using chloroacetic anhydride to facilitate thioether formation with a side chain thiol group on the linker. This necessitated the presence of a free thiol group on the ligand, therefore cysteine was chosen as the N-terminal amino acid of the linker. Conceivably, another thiol such as mercaptopropionic acid would have been less complicated because it does not have the amino group found in cysteine, but from an analytical perspective it is preferable to have the free amino group present as it more readily takes on a positive charge, which makes it easier to monitor synthesis of the compound by mass spectrometry.

A second modification to the Zhang *et al.* strategy was using an alternative method to cyclise the scaffold. Zhang's method uses on-resin cyclisation with an Alloc group to protect the C-terminus during synthesis (Figure 4.3). Previous experience of on-resin cyclisation reactions has shown that the degree of success is sequence specific, some peptides, such as  $\beta$ -amyloid 1-42 aggregate with neighbouring peptide chains during synthesis preventing further peptide synthesis, let alone on-resin cyclisation; this is in spite of the presence of large bulky protecting groups on the side chains of some amino acids in the resin-bound peptide and often the aggregation manifests itself as a visible reduction in the volume of the peptidyl-resin. Furthermore, on-resin cyclisation requires low resin substitution and the high cost of resin means that large scale syntheses are prohibitively expensive.



**Figure 4.3.** On-resin cyclisation using Allyl protecting group (Alloc) as used by Zhang *et al.*, 2004.

A number of possible ligation reactions were contemplated. The simplest way to replicate the work of Zhang *et al.*, without an on-resin cyclisation, would be to synthesise the peptide on a hyper acid-labile resin, cleave it as a protected peptide and attempt the head-to-tail cyclisation reaction in solution. It is often possible to achieve solution phase cyclisations because the higher probability of intramolecular collisions favours cyclisation over intermolecular polymerisation, particularly if a high dilution factor is used. However, the presence of protecting groups on the peptide make the chain less flexible and often polymerisation occurs anyway. Therefore, the solution-phase cyclisation method was discounted as an appropriate method. Often in peptide synthesis alternative methods of cyclisation cannot be contemplated because achieving a specific peptide sequence is a key part of the synthesis. However, in this case it was the architecture that was the overall aim and the building blocks used to achieve it were interchangeable.

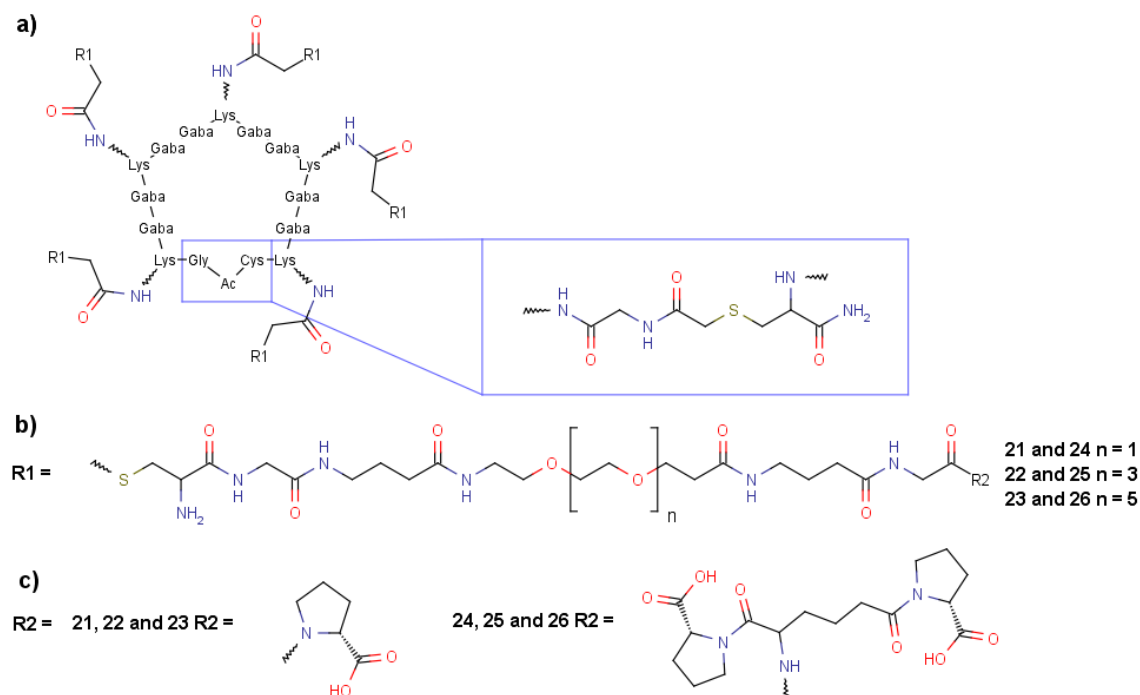


A second methodology used to cyclise peptides is native chemical ligation. In this method, an active thioester is formed at the C-terminus of a peptide that has an N-terminal cysteine. The thioester is formed after cleavage, from a suitable resin such as 2-Chlorotrityl resin, on a protected peptide using benzyl mercaptan (Dawson *et al.* 1997). The thioester species is rapidly displaced by the thiol side chain of cysteine, which subsequently rearranges to yield a peptide bond between the C-terminus and the N-terminal cysteine, which also regenerates the free thiol on the N-terminal cysteine. The disadvantage of thioester formation is the use of benzyl mercaptan which is highly malodorous. Due to this considerable drawback, thioester formation was discounted as a viable method, but it influenced the final decision to use cysteine as a key component of the cyclisation reaction. Literature research found previous use of thioether bonds to form cyclic peptides for multivalent ligands demonstrating that an intramolecular thioether cyclisation would be a realistic methodology (Jones *et al.*, 1998; Jones *et al.*, 1999). Therefore, it was decided that the thioether method used attach the linker to the central scaffold would also be applied to the intramolecular reaction.



**Figure 4.3.** General reaction scheme of thioether bond formation by thiol alkylation with an alkyl halide performed in aqueous solvent under mildly basic conditions.

The detailed design points discussed above led to the final chemical structure of the ligands pictured in Figure 4.4. The lengths of the linkers were estimated to be of sufficient length by generating the molecules using the ProdrG server (Schuettelkopf and Van Aalten, 2004) and measuring the molecule length in Coot (Emsley and Cowtan, 2004), while the size of the central scaffold was approximated based on the values from Zhang *et al.* (2004).



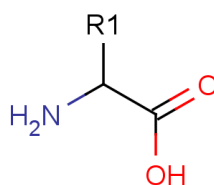
**Figure 4.4.** Chemical structure of each component for the proposed multivalent ligands for SAP and the combined structure. a) Central scaffold with thioether cyclisation. b) linkers with Cys-Gly-GABA-PEG<sub>n</sub>-GABA-Gly structure. c) headgroups, left: D-proline, right: amino-CPHPC.

#### 4.1.2. Solid Phase Peptide Synthesis

Solid Phase Peptide Synthesis (SPPS) is a widely used technique for synthesising peptides. In SPPS, a solid phase support, generally consisting of 1% cross-linked divinyl benzene, is functionalised with a linker, which is a functional group capable of forming a covalent attachment between the solid phase and an amino acid (Merrifield, 1963). Amino acids are attached to the solid phase through their C-terminal carboxyl group, leaving the N-terminal amino group free for covalent attachment of the next amino acid. The concentration (substitution) of linkers on the solid phase is varied to obtain a balance between yield and reaction efficiency. High concentrations of linker produce a higher peptide yield, but when peptide chains reach a length greater than 9 amino acids they can start to sterically hinder reactions on neighbouring peptide chains, leading to incomplete reactions that lower the purity and yield of the final product. Typically, a substitution of 0.4 mmol g<sup>-1</sup> achieves the appropriate balance.

Amino acids have the general structure shown in Figure 4.5. In SPPS, amide (peptide) bonds are formed between the amine of the resin-bound nascent peptide chain and the

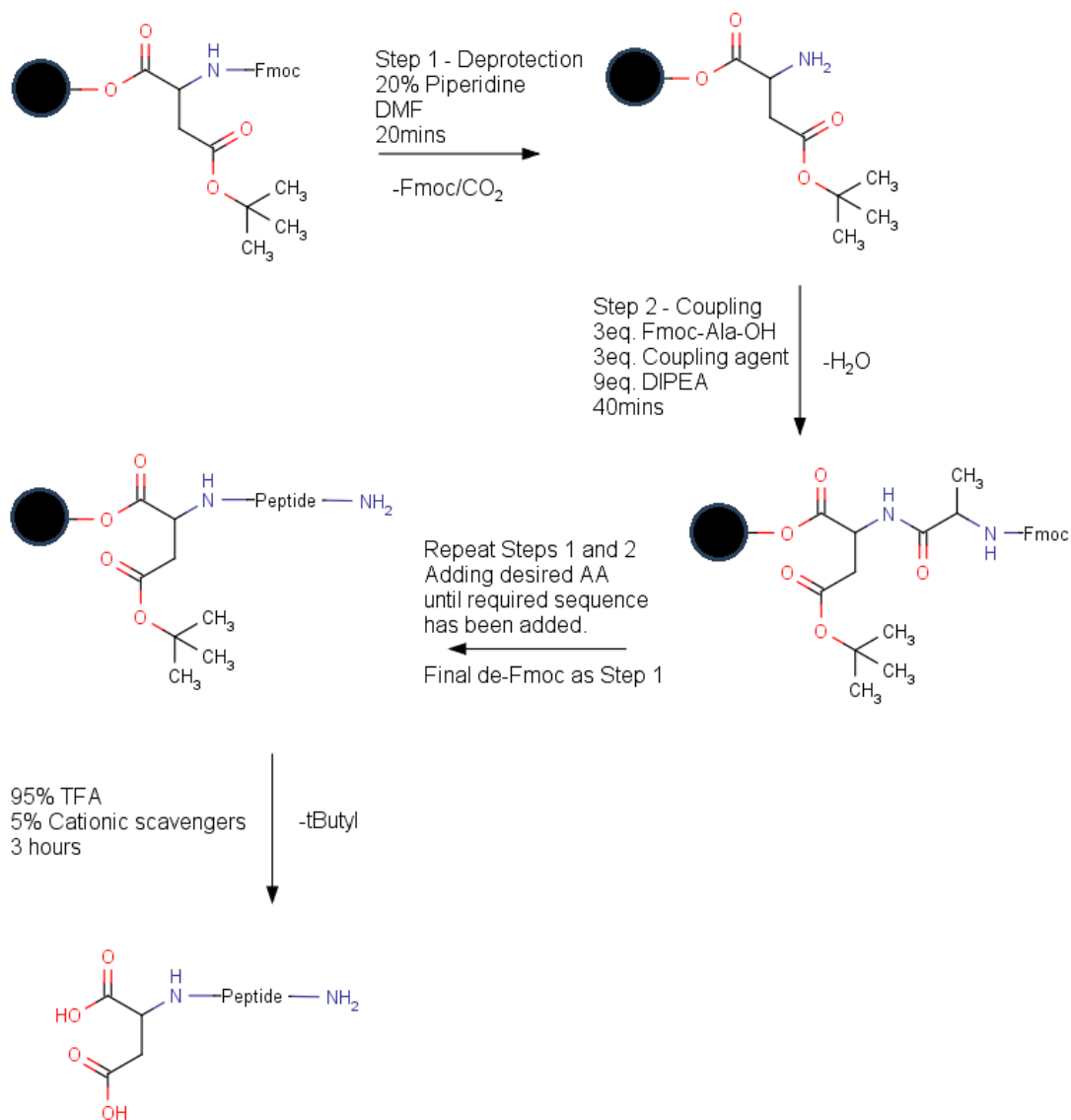
carboxyl group of a solution phase amino acid. There are a vast number of commercially available amino acids many with R-groups that would engage in undesirable side reactions in the reaction conditions used for peptide bond formation. SPPS employs a range of both side chain and N-terminal protecting groups to minimise side reactions and direct bond formation to the desired site. The two most common protecting group strategies in SPPS are 'Boc' and 'Fmoc', named after the structure of the N-terminal protecting group. Fmoc, Fluorenylmethoxycarbonyl, SPPS is more widely used because it does not require the use of dangerous hydrofluoric acid (HF) to liberate the peptide from the solid phase (Carpino and Han, 1972). Furthermore, Fmoc SPPS allows for 'orthogonal protection' in which the side chain protecting groups are fully stable to the conditions used to remove the N-terminal protecting group and *vice versa* (Barany and Merrifield, 1977).



**Figure 4.5.** Generic structure of an  $\alpha$ -amino acid with a primary amine.

The scheme in Figure 4.6 outlines the general processes of SPPS, a more substantial explanation can be found in *Solid Phase Synthesis: A Practical Approach* (Atherton *et al.*, 1989). In Fmoc SPPS, the C-terminus of a solution-phase amino acid is reacted with the N-terminus of a solid phase-bound amino acid, in the presence of a non-nucleophilic base, to form an amide (peptide) bond. Peptide bond formation does not happen spontaneously, so a coupling agent, such as Benzotriazole-1-yl-oxy-tris-(dimethylamino)-phosphonium hexafluorophosphate (BOP), is used to convert the C-terminus of the incoming amino acid to an active ester that readily reacts with the free amino group of the nascent peptide. A high excess of solution phase reagents relative to the solid phase amino acid ensures greater than 99% reaction completeness in 40 minutes. The excess reagents are then easily removed by repeated washing of the resin with an aprotic solvent such as dimethylformamide (DMF). The Fmoc group of the newly coupled amino acid is then removed by a 20 minute treatment with 20% Piperidine/DMF. After the resin is thoroughly washed with DMF, the process of amino

acid addition and deprotection is repeated until the desired peptide has been synthesised. Subsequent treatment of the resin-bound peptide with 95% trifluoroacetic acid and 5% cation scavengers (e.g. triisopropylsilane) liberates the peptide from the solid phase and removes any acid-labile side chain protecting groups. Purification by High Performance Liquid Chromatography (HPLC) yields a high purity peptide of the desired sequence.



**Figure 4.6.** Example scheme of the standard reaction steps of Fmoc solid phase peptide synthesis using preloaded Fmoc-Asp(OtBu) resin and coupling of Fmoc-Ala-OH. Step 1: Deprotection with 20% piperidine in DMF for 20 minutes. Step 2: Coupling of amino acid using Fmoc-protection on the N-terminus of the incoming amino acid, a coupling agent, such as BOP, to expedite the reaction and diisopropylethylamine (DIPEA) as the non-nucleophilic base. Steps 1 and 2 are repeated with the relevant Fmoc-protected amino acid until the desired sequence has been synthesised. Cleavage of the peptide from the resin and side chain deprotection is achieved simultaneously using 95% TFA, 5% cationic scavengers.

#### 4.1.2.2. Synthesis of non-standard peptides

Peptide chemistry is a highly malleable medium in which to design molecules with specific properties. The amide bonds provide an unreactive backbone, while the wide variety of available side chain R-groups gives almost limitless customisation of a peptide sequence and its properties. Addition of unnatural amino acids, such as  $\gamma$ -aminobutyric acid (GABA) or selenomethionine, is no more complicated than the inclusion of any standard amino acid. C-terminal amide formation can be achieved through the use of dedicated solid phase linkers such as Rink amide. N-terminal modifications can be performed both on-resin or in solution post-cleavage. The presence of side chain amino and thiol groups in amino acids, such as lysine and cysteine, allow for facile ligation of peptides to a host of other molecules. For these reasons, Fmoc SPPS is the ideal choice for the synthesis of multivalent ligands for SAP.

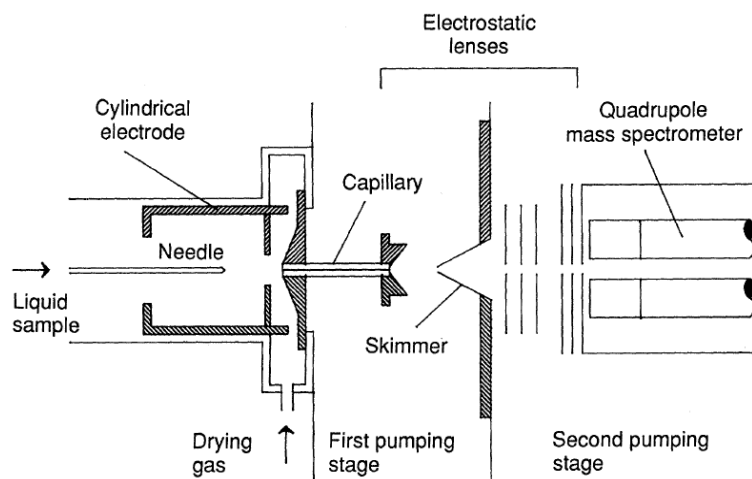
#### 4.2. Materials and Methods

Fmoc- $\gamma$ -aminobutyric acid (Fmoc-GABA-OH), chloroacetic anhydride, trifluoroacetic acid, triisopropylsilane, anisole, thioanisole and ammonium bicarbonate were purchased from Sigma Aldrich (Dorset, UK). 2-Chlorotrityl chloride resin was purchased from Iris Biotech (Marktredwitz, Germany). Fmoc-protected amino acids, coupling agents, H-D-Pro-OtBu, Rink Amide MBHA low loaded resin and various length Fmoc-PEG<sub>n</sub>-OH were purchased from Novabiochem (Merck KGaA; Darmstadt, Germany). Dimethyl sulfoxide (DMSO) and Diisopropylethylamine (DIPEA) were acquired from Alfa Aesar (Ward Hill, MA). Dimethylformamide (DMF) and dichloromethane (DCM) were purchased from AGTC Bioproducts (National Diagnostics; Hesse, UK).

Mass spectrometry (MS) was performed on a Micromass ZMD4000 Mass Spectrometer. Liquid chromatography mass spectrometry (LCMS) was performed on an Agilent 1100 series LCMS. Peptides were purified by reverse-phase high-performance liquid chromatography (HPLC) on a Gilson HPLC system with two 306 pumps, 118 UV/Vis detector, 806 manometric module and 811B dynamic mixer. Peptides were successfully purified to between 80-97% purity using an acetonitrile/water (0.1% v/v TFA) mobile phase and a Phenomenex Gemini 5 $\mu$  C18 110Å column as the solid phase.

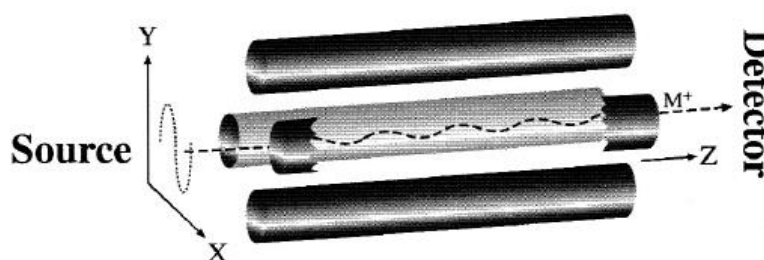
MS was used extensively in this chapter and the next for qualitative structure determination. Therefore, the basic theory of the technique will be briefly introduced. MS is widely used in organic chemistry and the principal technique for analysis during peptide synthesis. MS involves applying either a positive or negative charge to molecules in the sample of interest and passing the sample through a mass analyser, which separates ions based on their mass/charge ( $m/z$ ) ratio. Once separated ions arrive at the detector, which measures the abundance of different ions in a given mass range (Ho *et al.*, 2003).

There are a variety of methods used for ionising the sample, which can be classified as either “hard” or “soft”. Hard ionisation techniques, such as electron bombardment, tend to cause fragmentation of molecules and therefore are not typically applied to polypeptide analysis. Soft ionisation uses gentler methods to apply charge to sample molecules. Electrospray ionisation (ESI) is an example of soft ionisation MS and was used in this study for both assessing peptide synthesis reactions and multivalent ligand interactions with SAP. Electrospray ionisation uses a nebuliser needle with an electrified tip to apply charge to solvent droplets containing the molecule of interest (Figure 4.7). The sample is injected onto the apparatus in a small quantity of solvent, such as 1:1 acetonitrile:water solution. For positive ionisation, an acid such as 0.1% v/v trifluoroacetic acid is added to improve sample ionisation. The sample enters the cone of the spectrometer and is mixed with a drying gas, which forms an aerosol with the sample solvent. The droplets of the sample become charged as they leave the cone and are dried further by a second stream of drying gas flowing counter to the direction of the solvent. Coulombic explosions reduce the size still further until the molecules themselves are charged. The charged species is propelled into the mass analyser by repulsion from a cylindrical electrode surrounding the nebuliser needle.



**Figure 4.7.** Diagram of the sample nebuliser needle and mass analyser of a typical electrospray mass spectrometer (ESI-MS). Adapted from Fenn *et al.*, 1989.

The mass analyser generally consists of either 4 or 6 parallel rods (quadrupole or hexapole) arranged radially to which alternating charges are applied with a specific frequency. As the ions pass through the centre of the analyser the ions are attracted and repelled by the alternating charges of the poles, resulting in a unique undulating trajectory (Figure 4.8). Ions with a high  $m/z$  ratio change direction rapidly and eventually spin out of the analyser or collide with the poles. Heavier ions with low  $m/z$  ratios change direction more slowly and as a result collide with the poles. Only ions of the correct  $m/z$  for a particular frequency will pass through the analyser and reach the detector. Ions that collide with the detector result in an emission of electrons, which are liberated from the surface of the detector and enter the electron multiplier horn. The electron multiplier horn amplifies the electrical signal such that it can be recorded as a peak on a mass spectrum. The frequency at which the charge is flipped in the mass analyser poles determines which ions will reach the detector; varying this through the lifetime of a MS analysis gives a full spectrum for the desired mass range.



**Figure 4.8.** Diagram of the Quadrupole and the path of ions through the mass analyser. Adapted from Ho *et al.*, 2003.

#### 4.2.1. Synthesis of peptides

All peptides were synthesised by Fmoc SPPS as described by Atherton *et al.* (1989), non-standard modifications are detailed in the following sections. The multivalent ligands were synthesised in several fractions, the central cyclic scaffold **5**, the pentameric linker-headgroups (**6**, **7**, **8**), the decameric linkers (**9**, **10**, **11**) and the decameric headgroup (**14**). The fractions were then combined into the desired multivalent product by a series of solution phase reactions.

#### 4.2.2. Small scale cleavage of peptides for mass spectrum monitoring during synthesis

The progress of SPPS can be monitored between reaction steps by placing a small number of solid phase beads (typically 5-10) in a small test tube adding 100 $\mu$ l of TFA and 1 $\mu$ l of triisopropylsilane and heating the sample in a heating block at 100°C until all the TFA has evaporated. This method liberates sufficient peptide from the resin for MS analysis, which is prepared for injection by dissolving in 200 $\mu$ l of 1:1 water and acetonitrile solution. The sample is transferred to a microcentrifuge tube, briefly spun down and 10 $\mu$ l of the supernatant injected onto the mass spectrometer.

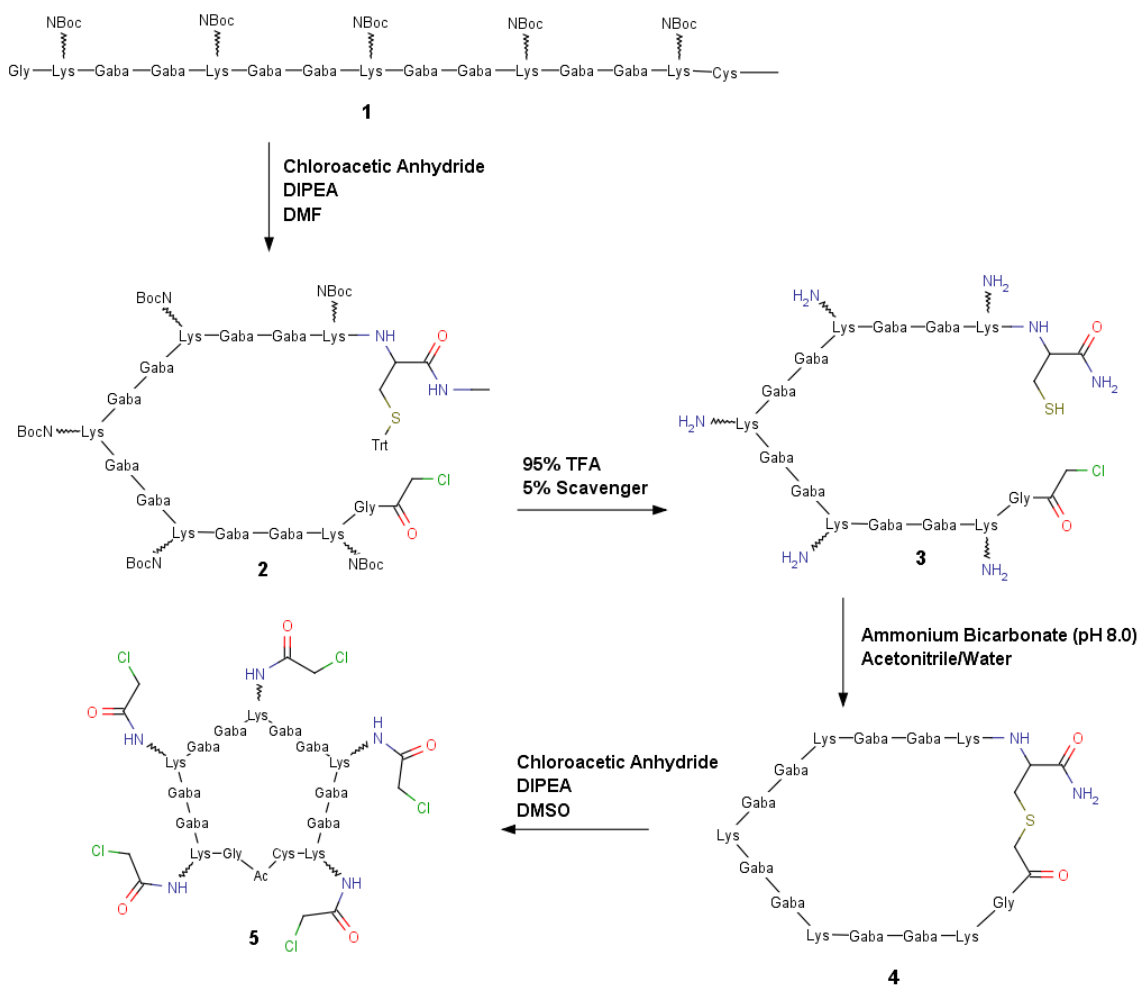


### 4.2.3. Synthesis of central scaffold 5

#### 4.2.3.1. SPPS of uncyclised central scaffold 1

Linear peptide **1** (Figure 4.9) was synthesised on low-loaded ( $0.45\text{mmol g}^{-1}$ ) Rink amide MHBA resin. Rink resin yields a peptide with a C-terminal amide upon cleavage with 95% TFA; this can be a useful way of reducing the reactivity of a carboxylic acid and preventing undesirable side reactions during post-cleavage modifications. MS analysis was performed to confirm the identity of the desired product and assess reaction completeness before every Fmoc-deprotection and prior to coupling of the next Fmoc-protected amino acid. The mass spectrum of the crude resin-bound product **1** showed the following peaks: M+2H ion: 751.0 m/z, relative abundance (r.a.): 75%, M+2H ion + C-terminal phenol: 803.7 m/z, M+3H ion: 501.2 m/z, r.a.: 100%, M+3H ion + C-terminal phenol: 536.5 m/z, r.a.: 70%. +106 peaks were observed, indicating a portion of the Rink amide linker had cleaved off the resin attached to the peptide, this is commonly known to be an issue with this linker. It only presented itself in this mass spectrum because of the elevated temperature of the cleavage conditions used to analyse peptides during SPPS (section 4.2.3); no +106 was observed for the final room temperature cleavage.

## Chapter 4. Synthesis of multivalent ligands for SAP

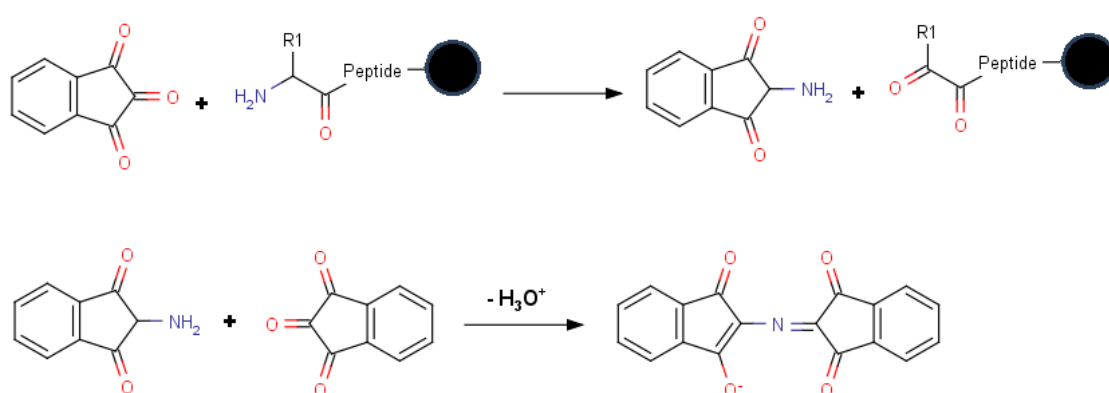


**Figure 4.9.** Synthesis of central cyclic scaffold 5.

### 4.2.3.2. Formation of 2 by addition of chloroacetyl group to the N-terminus of 1

Resin-bound peptide **1** was reacted with 3 molar equivalents (equiv.) chloroacetic anhydride dissolved in 2ml DMF with 9 equiv. DIPEA over 30 minutes. The unreacted material was then removed from the peptidyl resin by filtering and the resin washed three times with DMF. The extent of the reaction was determined by MS and Kaiser test. The Kaiser test uses 2,2-dihydroxyindane-1,3-dione (ninhydrin) to test for the presence of primary amines, making it a simple, but crude, method of testing for reaction completeness during SPPS (Figure 4.10). A small sample of peptidyl-resin is placed in a small test tube containing 20µl ninhydrin/methanol solution and 40µl water containing potassium cyanide and phenol. The mixture is then heated for 3 minutes and the colour of the mixture observed. If a free primary amine is present then a “positive” dark blue/purple colour is observed, otherwise a negative yellow/straw colour is

observed. A Kaiser test of the chloroacetylated peptidyl resin gave a negative result and MS confirmed that the coupling reaction was complete. The resin was washed thoroughly with DCM to remove any remaining DMF prior to cleaving the peptide from the solid phase. For example: 0.6g of **1** reacted with 0.023g (0.135mmol) chloroacetic anhydride in 2ml DMF with 125 $\mu$ l (0.73mmol) DIPEA produced a crude mass spectrum of product **2**: M+2H ion: 788.9 m/z, r.a.: 20%, M+3H ion: 526.5 m/z, r.a.: 55%, M+4H ion: 395.2 m/z, r.a.: 100%, M+5H ion: 316.5 m/z, r.a.: 40% (product not isolated prior to next reaction, thus not quantified).



**Figure 4.10.** Reaction of ninhydrin with N-terminus of a nascent peptide chain, followed by reaction of the intermediate with ninhydrin to produce the chromophore responsible for the observed purple colour in a 'positive' result.

#### 4.2.3.3. Cleavage of **2** from the solid phase and purification of product **3**

To break linkages between the peptide and the solid phase and to remove all acid labile protecting groups, dry peptidyl-resin was gently stirred for 3 hours in a solution of 90-95% TFA, 5-10% cationic scavengers. The cationic scavengers consisted of equal volumes of triisopropylsilane, water, anisole and thioanisole. After 3 hours, the cleavage solution was separated from the resin by gravity filtration assisted by a small application of positive pressure and the resin washed thrice with 2ml 100% TFA. The 6ml TFA used to wash the resin was pooled with the rest of the cleavage cocktail. Typical reaction quantities: 0.6g **2** was mixed with 6ml TFA and 600 $\mu$ l scavenger for 3 hours.

#### 4.2.3.4. Removal of TFA from crude **3**

Ether precipitation was used to separate peptides from the concentrated TFA if their sequence did not contain PEG. In this method, the cleavage cocktail is pooled in a 40ml centrifuge tube. 20ml diethyl ether (ether) is then added, causing a thick white precipitate to form. After thorough mixing, the solution is placed in a centrifuge at 3200 rpm for two minutes. The supernatant is removed, another 20ml ether added and the precipitate agitated with a metal spatula to liberate TFA trapped within the precipitate. The sample is returned to the centrifuge for a further two minutes. The supernatant is removed once more and 5ml water added to the precipitate. The solid is agitated before freeze-drying. After freeze-drying the crude peptide material was purified by HPLC utilising a 1-21% acetonitrile gradient over 40mins. The product eluted after 11mins, MS revealed the product to be free from peptidyl impurities. Purified mass spectrum of product **3**: M+2H ion: 788.9 m/z, r.a.: 25%, M+3H ion: 526.5 m/z, r.a.: 100%, M+4H ion: 395.2 m/z, r.a.: 75%. Product not isolated before next reaction.

#### 4.2.3.5. Formation of **4** by thioether alkylation

The cyclisation of **3** was achieved immediately after HPLC purification. The pH of the aqueous sample eluted from the HPLC was raised to between 7.5-8.0 using ammonium bicarbonate and stirred for 3 hours open to the atmosphere. Under basic conditions the chloride in the chloroacetyl group undergoes nucleophilic substitution with the thiol of the C-terminal cysteine resulting in cyclic peptide **4**. After 3 hours, a 10µl sample was acidified with formic acid and analysed by MS, which showed the reaction to be complete. The reaction mixture was freeze-dried and purified by HPLC. Typical results of cyclisation of peptide synthesised on 2x 600mg Rink resin: Yield: 61mg (43.9%). Mass spectrum of **4**: M+2H ion: 771.1 m/z, r.a.: 10%, M+2H ion + trifluoroacetate: 828.1 m/z, r.a.: 10%, M+3H ion: 514.4 m/z, r.a.: 100%, M+3H ion + trifluoroacetate: 552.5 m/z, r.a.: 20%.

#### **4.2.3.6. Formation of 5 by reaction of chloroacetic anhydride with the lysine side chains of 4**

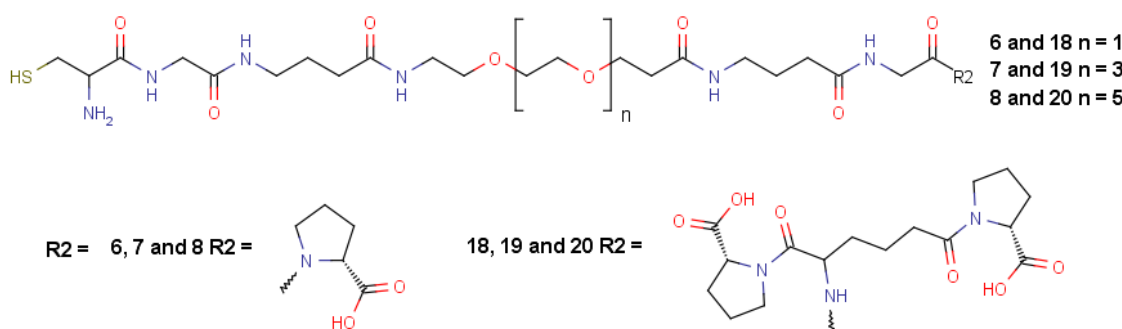
The five lysine side chains of purified **4** were acetylated with chloroacetic anhydride in DMSO under basic conditions. The reaction was determined complete by MS after 20mins and immediately purified by HPLC. For example 50mg (0.032mmol) **4** was dissolved in 2ml DMSO with 78mg chloroacetic anhydride and 125 $\mu$ l (0.72mmol) DIPEA added and the mixture stirred for 20mins. After 20mins no **4** was detectable by MS, the product was treated with 100 $\mu$ l formic acid and immediately purified by HPLC running a gradient of 5-95% acetonitrile over 40mins, the product eluted after 20mins. MS of purified **5**: M+2H ion: 961.4 m/z r.a.: 100%, unrelated peaks contaminating spectrum: 373.6 m/z r.a.: 80%, 295.5 m/z r.a.: 40%. Product not isolated before next reaction.

#### **4.2.4. SPPS of 6, 7 and 8**

##### **4.2.4.1. Coupling of first amino acid to 2-chlorotrityl chloride resin**

Coupling the first amino acid to 2-Ctr resin involves formation of a stable ester bond by nucleophilic substitution of the chloro group on the trityl linker of the resin by the C-terminal carboxyl group of an Fmoc-protected amino acid. 2-Ctr resin is often used in the synthesis of peptides with a C-terminal proline because of their propensity to undergo diketopiperazine formation during SPPS. Diketopiperazine formation occurs after addition of the second amino acid (Gisin and Merrifield, 1972). Once deprotected, some dipeptides have a tendency to arch back towards the resin and attack the ester linkage between the peptide and the solid phase. The peptide is liberated from the solid phase as a stable 6-member heterocycle and a free hydroxyl group forms on the linker (Figure 4.16). Upon addition of the next amino acid, a new ester linkage forms between the incoming amino acid and the hydroxyl group on the resin. As the synthesis continues it becomes apparent by MS that a proportion of the peptide is missing the first two amino acids. Quite often it is not possible to remove the impurity by HPLC, but the extent of the side reaction can be suppressed by using a solid phase linker, such as 2-Ctr, that has large bulky groups to sterically hinder attack of the peptide ester linkage.

The desired product peptides, **6**, **7** and **8**, (Figure 4.11) were short enough that highly substituted resin could be used without the risk of detrimentally affecting the quality of the synthesis. Therefore in a typical synthesis, 0.6g 2-Ctr resin was reacted with 0.3mmol Fmoc-protected amino acid in 50:50 DMF:DCM with 375 $\mu$ l DIPEA for 1 hour, on the assumption that 1/3 of the amino acid would successfully couple to the solid phase. After 1 hour the liquid was removed by filtering the resin and flushing with 3ml DMF. The resin was then treated with 1ml methanol in 3ml DMF with 375 $\mu$ l DIPEA to cap any unreacted linkers. After capping and washing the resin, peptide elongation was performed by standard SPPS and the peptide subsequently cleaved by the same procedure as for **3** (section 4.2.4.3). However, due to the presence of PEG in the sequence of peptides **6**, **7** and **8**, a different method was used to remove the TFA from the crude peptide after cleavage.



**Figure 4.11.** Structure of linker-headgroups **6**, **7**, **8**, **18**, **19** and **20**.

#### 4.2.4.2. Separation of peptides containing polyethylene glycol from TFA

PEG can increase the solubility of peptides in ether, preventing effective precipitation. Therefore, the TFA is removed by rotary evaporation. The sample is placed on the rotary evaporator until reduced to a viscous oil. 5ml water is then added to the sample and freeze-dried.

#### 4.2.4.3. Purification of **6**, **7** and **8**

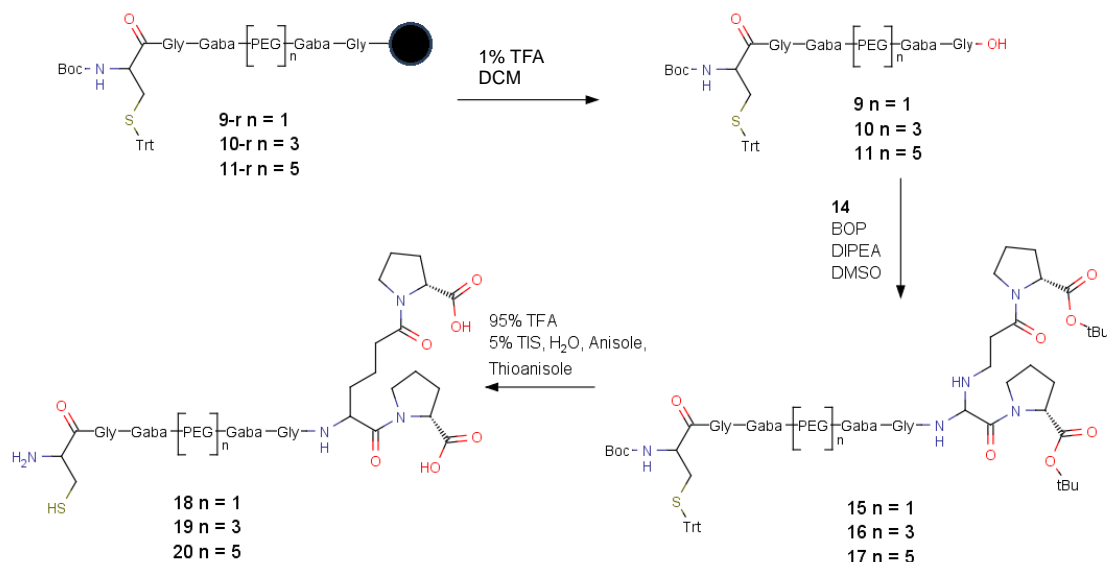
HPLC purification was performed on crude peptides **6**, **7**, **8** using a 1-51% acetonitrile gradient over 40mins the following retention times were recorded: **6** eluted after 18mins, **7** after 21mins, **8** after 23mins. The purified mass spectra for each component were as follows: **6**: M+1H ion: 648.6 m/z, r.a.: 100%. **7**: M+1H ion: 750.4 m/z, r.a.:

100%, M+2H ion: 376.1 m/z, r.a.: 40%. **8**: M+1H ion: 848.5 m/z, r.a.: 100%, M+2H ion: 420.0 m/z, r.a.: 40%. Yield (2x 600mg syntheses of each product): **6**: 181mg (23.3%), **7**: 107mg, (11.9%), **8**: 140mg (13.7%). Percentage yields based on 1/3 of first amino acid coupling to solid phase.

#### 4.2.5. Synthesis of **18**, **19** and **20**

##### 4.2.5.1. SPPS of **9**, **10** and **11**

Unlike the pentavalent linker-headgroups, the linkers and the headgroups for the decavalent compounds were synthesised separately and then combined by solution phase reaction (Figure 4.12). Linkers **9**, **10** and **11** were synthesised on 2-Ctr resin because the triphenyl structure of the 2-Ctr linker allows the ester bond to be broken under weakly acidic conditions, so peptides can be cleaved from the resin without removing protecting groups. Protected peptides can then be modified further in solution post-cleavage. The N-terminal cysteine was coupled as Boc-Cys(Trt)-OH because it negates the need for a final deprotection step and, in the case of **9**, **10** and **11**, allows the peptide to be cleaved with N-terminal protection, permitting a solution phase coupling reaction at the C-terminus of the peptide without the risk of polymerisation.



**Figure 4.12.** Synthesis of decameric linker headgroups **18**, **19** and **20**.

#### 4.2.5.2. Cleaving protected peptides **9**, **10** and **11** from 2-Ctr resin using 1% TFA

**9**, **10** and **11** required further post-cleavage modification; therefore they were cleaved by repeatedly treating the resin with 2ml 1% TFA in DCM to remove the peptide from the hyper-acid labile 2-Ctr resin without cleaving the protecting groups. After cleaving the peptide with a total volume of 20ml 1% v/v TFA/DCM, the solvent was pooled and separated from the crude peptide by rotary evaporation and then freeze-dried. For example, compound **9**: M+1H ion: 893.5 m/z, r.a.: 100%, M+1H ion -Trityl protecting group fragmentation: 651.4 m/z, r.a.: 15%.

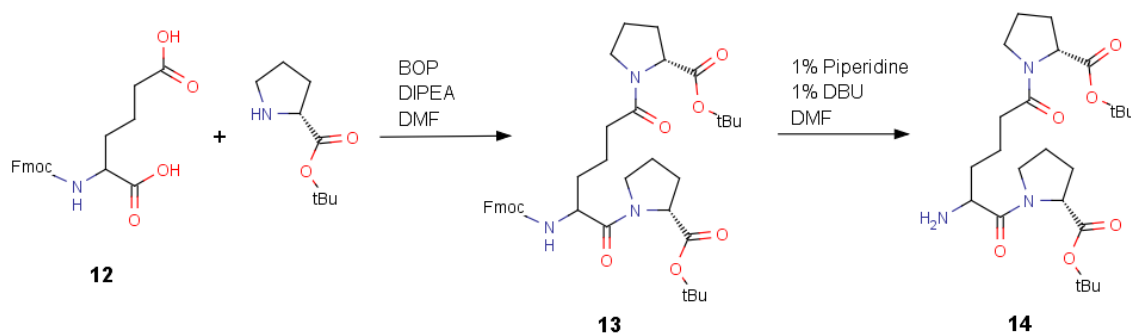
#### 4.2.5.3. Synthesis of **14**

hGlu was Fmoc-protected by dissolving the amino acid in water with 4 equiv. sodium bicarbonate. 1.2 equiv. Fmoc-succinimide (Fmoc-OSu) was dissolved in dioxane and added dropwise to the aqueous solution. After stirring overnight, the dioxane was removed by rotary evaporation to leave a turbid aqueous solution. Acidification of the aqueous solution with a couple of drops of concentrated HCl caused precipitation of a dense white solid. The precipitate was separated from the aqueous phase by vacuum filtration and dried on a dome freeze-dryer. For example: 0.1g (0.62mmol) hGlu and 0.2g (2.38mmol) Sodium Bicarbonate were dissolved in 4ml water. 0.25g (0.74mmol) Fmoc-OSu in 5ml dioxane was added dropwise over two hours. Crude yield: 126mg (60.1%) Mass spectrum of crude Fmoc-hGlu-OH (**12**): M+1/2H ion (product flying as dimer on MS): 767.4 m/z, r.a.: 80%, M+1/2H +NH<sub>3</sub>: 784.4 m/z, r.a.: 35%, M+1H ion: 384.4 m/z, r.a.: 100%, M+1H ion +NH<sub>3</sub>: 401.3 m/z, r.a.: 30%. Unidentified contaminating peak: 206.3 m/z, r.a.: 60%. Ammonia contamination was observed because the MS used is open access and used for characterisation of products purified under basic conditions, so occasional contamination is to be expected.

1 equiv. of **12** was then reacted in solution with 2 equiv. of H-D-Pro-OtBu using 2 equiv. benzotriazole-1-yl-oxy-tris-(dimethylamino)-phosphonium hexafluorophosphate (BOP) as the coupling agent and 6 equiv. DIPEA (Figure 4.13). The reaction mixture was acidified with 200µl formic acid and purified by HPLC running an acetonitrile gradient from 40-100% over 40mins. The desired product, Fmoc-hGlu(dPro-OtBu)-dPro-OtBu (**13**) was eluted after 25mins free from any remaining starting material and



was immediately freeze-dried. For example: 90mg (0.23mmol) **12** and 107mg (0.63mmol) H-D-proline-OtBu and 225mg (0.51mmol) BOP were dissolved in 2ml DMSO with 250 $\mu$ l (1.47mmol) DIPEA and stirred for 1 hour. Yield of **13**: 73mg (44.9%). MS of **13**: M+1H ion: 690.7 m/z r.a.: 100%.



**Figure 4.13.** Decameric headgroup synthesis.

Once freeze-dried, the Fmoc group was removed from **13** by dissolving the pure starting material in 2% piperidine, 2% 1,8-Diazabicycloundec-7-ene (DBU) in DMF and stirring for 2mins. After 2mins, MS analysis showed complete N-terminal deprotection of **13**. 100 $\mu$ l formic acid was added and the mixture immediately injected onto the HPLC running an acetonitrile gradient between 15-95% over 40mins, the desired product, hGlu(dPro-OtBu)-dPro-OtBu (**14**) was eluted after 18mins free from peptidyl impurities and placed on the freeze-dryer. For example: 73mg **13** was dissolved in 2ml 2% piperidine, 2% DBU DMF giving a purified yield of 35mg (49.4%) of **14**. MS of **14**: M+1H ion: 468.7 m/z r.a.: 100%, M+1/2H ion: 935.9 m/z, r.a.: 60%.

#### 4.2.5.4. Fragment condensation between decameric linkers (**9**, **10** and **11**) and **14**

1 equiv. **14** was ligated to 1.2-2 equiv. of decameric linker **9**, **10** or **11** by solution phase reaction in 10ml DCM with 1.2 equiv. BOP as the coupling agent and 3 equiv. DIPEA as the base (Figure 4.12). After mixing for 1 hour, a 10 $\mu$ l sample was taken and the DCM removed in a heating block at 100°C. The dried sample was treated with 2 $\mu$ l triisopropylsilane and 100 $\mu$ l TFA and placed in a heating block until all the acid had evaporated. The residue was dissolved in 200 $\mu$ l 50:50 acetonitrile:water and analysed by MS, which showed no detectable **14**, indicating that the reaction was complete. The DCM was removed by rotary evaporation, redissolved in 3ml DMSO, acidified with

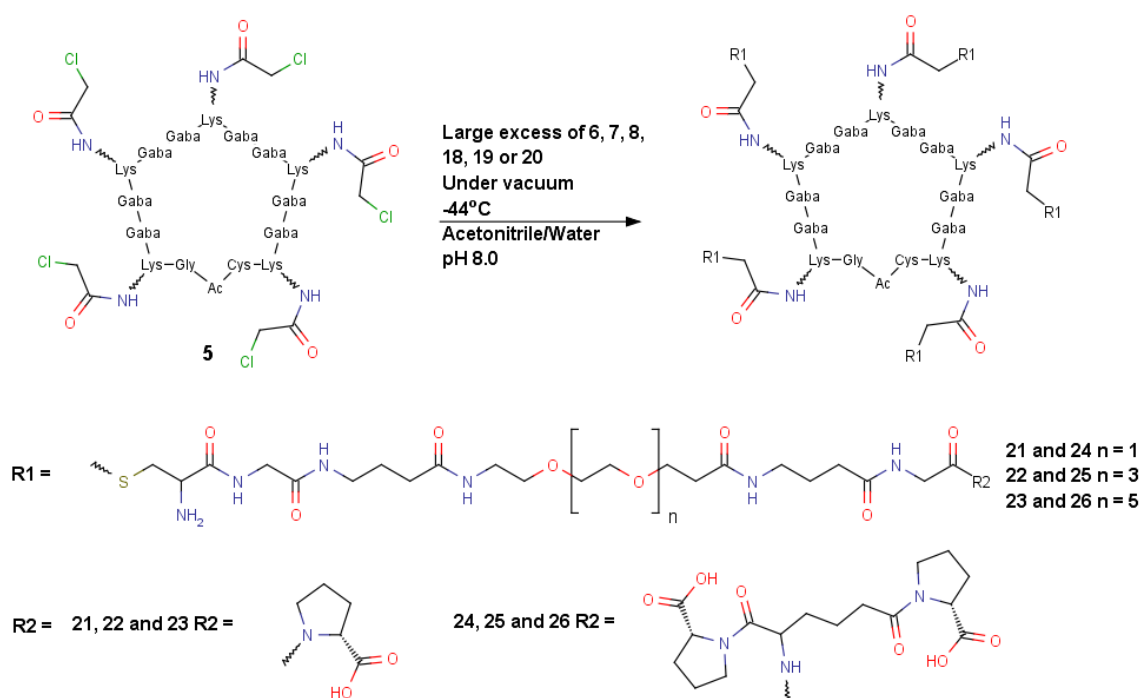
100 $\mu$ l formic acid and purified by HPLC running an acetonitrile gradient of 15-95% over 40mins. The desired products eluted after 34mins free from all peptidyl impurities. For example: 35mg (0.07mmol) **14** reacted with 75mg (0.08mmol) **9** with 37mg (0.08mmol) BOP and 36 $\mu$ l (0.21mmol) DIPEA in 7ml DCM gave the purified mass spectrum: M+1H ion: 1342.9 m/z, r.a.: 100%, M+1H ion -Trt fragmentation: 1100.9 m/z, r.a.: 15%, unidentified contaminant peak: 691.4 m/z, r.a.: 15%.

After freeze-drying the protected peptides were weighed, in total 45mg **15**, 115mg **16** and 170mg **17** was synthesised, as the products would not be isolated in solid form again before addition to the cyclic scaffold these values would serve as a basis for calculating expected yields of the multivalent products. The purified products were deprotected as in section 4.2.5 and then purified by HPLC using a gradient from 15 to 45% acetonitrile over 40mins. The desired products eluted after 18mins (**18**) and 21mins (**19**, **20**). The purified products were identified using MS and then freeze-dried. MS of deprotected, purified products: **18**: M+2H ion: 445.1 m/z r.a.: 100%; M+1H ion: 888.5 m/z r.a.: 75%. **19**: M+2H ion: 496.2 m/z, r.a.: 100%; M+1H ion: 990.9 m/z, r.a.: 70%; M+2H ion + K<sup>+</sup>: 515.3 m/z, r.a.: 15%. **20**: M+2H ion: 540.2 m/z, r.a.: 100%; M+1H ion: 1078.8 m/z, r.a.: 40%; M+2H ion +Na<sup>+</sup>: 551.3 m/z, r.a.: 15%.

#### 4.2.6. Formation of multivalent ligands

##### 4.2.6.1. Preparation of **21**, **22**, **23**, **24**, **25** and **26** by thioether formation between **5** and linker-headgroups **6**, **7**, **8**, **18**, **19** and **20**

**5** and one of **6**, **7**, **8**, **18**, **19** or **20** linker-headgroups (LH) were taken in solvent directly from the HPLC and mixed together (Figure 4.14). The LH was targeted to be in a greater than 15-fold molar excess to central scaffold, i.e. a 3-fold excess of LH to each reaction site. Once mixed the pH of the solution was raised to between 7.5-8.0 with ammonium bicarbonate. The reaction was immediately frozen and placed under vacuum on a dome freeze-dryer to reduce the rate of LH dimerisation by disulphide bond formation, a process requiring an oxidant such as molecular oxygen.



**Figure 4.14.** Multivalent ligand assembly, thioether formation reaction between central scaffold **5** and linker headgroups **6, 7, 8, 18, 19** and **20**.

The sample was slowly dried under vacuum and consequently yielded a partially reacted mixture of S-nLH, where S is scaffold **5** substituted with n LH joined to the scaffold through thioether bonds. The desired multivalent product, S-5LH was the major constituent, with S-4LH and S-3LH as the only multivalent impurities detectable, an example of this is shown by the mass spectrum of **22** in Figure 4.20. MS analysis also showed the unreacted LH was not significantly dimerised. Therefore, the sample was redissolved in 50:50 acetonitrile/water, pH 7.5-8.0 and once again placed under vacuum as described above. This process was repeated until S-5HL was the only detectable multivalent species in the mixture or until there was no significant improvement in the ratio of S-5HL to S-(5-m)HL achieved by further reaction on the freeze-dryer, where m is the number of ligands missing from a multivalent scaffold. At this point the dried sample was purified by HPLC.

#### 4.2.6.2. Purification of multivalent ligands and ligand-headgroups

Adequate separation was achieved using a gradient of AcN from 5-45% over 40mins. Samples were dissolved in 3ml water or DMSO and acidified using 100 $\mu$ l formic acid prior to injection. Mass spectra of the final purified products: **21**: M+3H ion: 1660.0 m/z, r.a.: 20%, M+4H ion: 1245.3 m/z, r.a.: 100%, M+4H ion + K<sup>+</sup>: 1254.8 m/z, r.a.: 15%, M+5H ion: 996.5 m/z, r.a.: 70%. **22**: M+4H ion: 1373.1 m/z, r.a.: 25%, M+5H ion: 1098.7 m/z, r.a.: 100%, M+6H ion: 915.8 m/z, r.a.: 35%. **23**: M+4H ion: 1483.2 m/z, r.a.: 75%, M+5H ion: 1186.8 m/z, r.a.: 100%, M+6H ion: 989.2 m/z, r.a.: 50%. **24**: M+4H ion: 1545.7 m/z, r.a.: 90%, M+4H ion +K<sup>+</sup>: 1555.3 m/z, r.a.: 10%, M+4H ion - 1x **18** + Cl: 1332.8 m/z, r.a.: 15%, M+5H ion: 1236.8 m/z, r.a.: 100%, M+5H ion -1x **18** + Cl: 1066.6 m/z, r.a.: 10%, M+6H ion: 1031.0 m/z, r.a.: 20%. **25**: M+4H ion: 1673.7 m/z, r.a.: 10%, M+5H ion: 1339.2 m/z, r.a.: 90%, M+6H ion: 1116.2 m/z, r.a.: 100%. **26**: M+5H ion: 1427.2 m/z, r.a.: 30%, M+6H ion: 1189.6 m/z, r.a.: 100%, M+7H ion: 1019.8 m/z, r.a.: 5%. Yield of final products: **21**: 20mg, (49.5%). **22**: 41mg, (72%). **23**: 43mg, (66%). **24**: 24mg, (48%). **25**: 21mg, (52%). **26**: 32mg, (56%). Yields calculated based on initial mass of **4** activated with chloroacetic anhydride for reaction with LH. See appendix for purified mass spectra of products.

#### 4.2.6.3. Preparation of 21- 26 as chloride salts

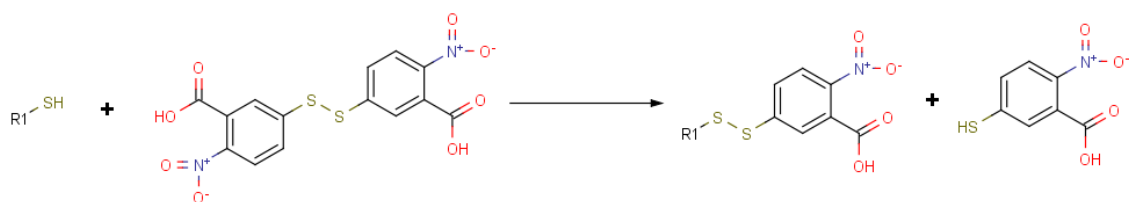
HPLC purification of peptides is generally performed with 0.1% TFA added to the mobile phase to improve peak resolution. Peptides are, therefore, present as trifluoroacetate salts after purification (Fields *et al.*, 1993). Trifluoroacetate ions have the potential to bind to SAP in a calcium-dependent manner through their carboxyl group, so it is essential to exchange this counter ion with another counter ion, such as chloride. This was achieved by thrice redissolving the freeze-dried product in dilute HCl (0.1M) and freeze-drying the solution. No attempt was made to confirm the removal of trifluoroacetate ions from the purified multivalent ligands, but evidence exists in the literature that a single freeze-drying with dil. HCl is sufficient for complete exchange with chloride ions (Andrushchenko *et al.*, 2007).

#### 4.2.6.4. Packaging and storage of multivalent ligands

Multivalent ligands dissolved in a small quantity of dilute HCl were aliquoted equally into 3 identical vials and a fourth containing 1mg and sealed under nitrogen. The 1mg vial was dispatched for amino acid analysis at the Protein and Nucleic Acid Chemistry Facility (University of Cambridge, UK) to determine the peptide content within the 1mg sample. The exact calculations were required for later determination of binding affinity to SAP by isothermal titration calorimetry discussed in Chapter 5.

#### 4.2.7. Ellman test for free thiol groups

Ellman's reagent was used to test for free thiol groups in the purified products of the intramolecular and intermolecular thioether formation reactions. To prepare the solution for thiol detection, 4mg Ellman's reagent, dithio-nitrobenzoic acid, is dissolved in a 1ml solution of 0.1M sodium phosphate pH 8.0. 10 $\mu$ l of this solution is mixed with small quantity of the purified material of interest. The sample is incubated for 15 minutes at room temperature and the colour observed. If a free thiol is present the solution will turn a clear yellow colour, otherwise the solution will remain colourless. No colour change was observed for the products of the intramolecular and intermolecular thioether reactions (**5**, **21**, **22**, **23**, **24**, **25** and **26**). However, a clear yellow colour was observed for the unreacted linkers **6**, **7**, **8**, **18**, **19** and **20**.



**Figure 4.15.** Reaction to test for free thiol groups using Ellman's reagent. Reduced thiol component responsible for the yellow colour observed for a positive result that indicates presence of a free thiol in the sample being analysed.

### 4.3. Results

#### 4.3.1. Final product yields

Table 4.1 summarises the yields and percentage yields of the multivalent ligands calculated from the original syntheses of the central cyclic scaffold **4** through to their final ligation to the linker-headgroups (LH). The yields range between 21-32%, which is in line with yields reported by Zhang *et al.* (2004). Crucially, the quantity produced is large enough for a variety of different investigations of multivalent binding with SAP.

Compound Number	Yield (mg)	Yield (%)
<b>21</b>	20	21.7
<b>22</b>	41	31.6
<b>23</b>	43	29.0
<b>24</b>	24	21.1
<b>25</b>	21	22.8
<b>26</b>	32	24.6

**Table 4.1.** Yield of multivalent ligands calculated by weighing mass of chloride salt of **21-26** after freeze-drying.

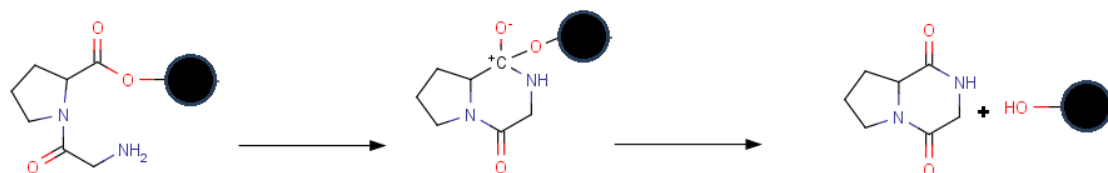
#### 4.3.2. Purity of 21-26

HPLC purification of the multivalent ligands was sufficient to obtain all but **21** (85%) and **22** (81%) at greater than 90% purity as determined by LCMS (Table 4.2). The LCMS data from which the purity for each ligand was calculated can be found in the appendix (section A.1).

Compound	Purity	Compound	Purity
<b>21</b>	85%	<b>24</b>	90%
<b>22</b>	81%	<b>25</b>	91%
<b>23</b>	96%	<b>26</b>	91%

**Table 4.2.** Purity of multivalent ligands indicated by LCMS.

Inspection of the mass spectra for pentavalent ligands **21-23** shows minor peaks equivalent to -154 m/z of the molecular ion, which is likely to be the species reducing the overall purity of the ligands (see appendix section A.2). This corresponds to one multivalent arm missing Gly-Pro from the C-terminus, which may have occurred by diketopiperazine formation during synthesis (Figure 4.16). Another explanation for the missing residues is that the coupling reaction did not proceed to completeness, but this can be discounted because there is no sign of the ligands missing just proline or just glycine; the pair of amino acids are missing simultaneously, indicating diketopiperazine (Gisin and Merrifield, 1972).



**Figure 4.16.** Diketopiperazine formation between glycine and proline during SPPS under basic conditions.

In this case, the ratio of 5 LH arms per scaffold has exacerbated the abundance of diketopiperazine afflicted arms (LH -GP) in the final molecule. If the abundance of LH -GP in the purified arms prior to ligation to **5** is just 2% then we would expect 10% of final product to contain 1x LH-GP.

No diketopiperazine formation was detected for the decavalent ligands because the synthetic strategy was different, the headgroup, **14**, was synthesised by solution phase reaction and although the decavalent linkers **9**, **10** and **11** were synthesised on the solid phase, the second amino acid was GABA, which has a backbone chain that is too long for diketopiperazine to form.

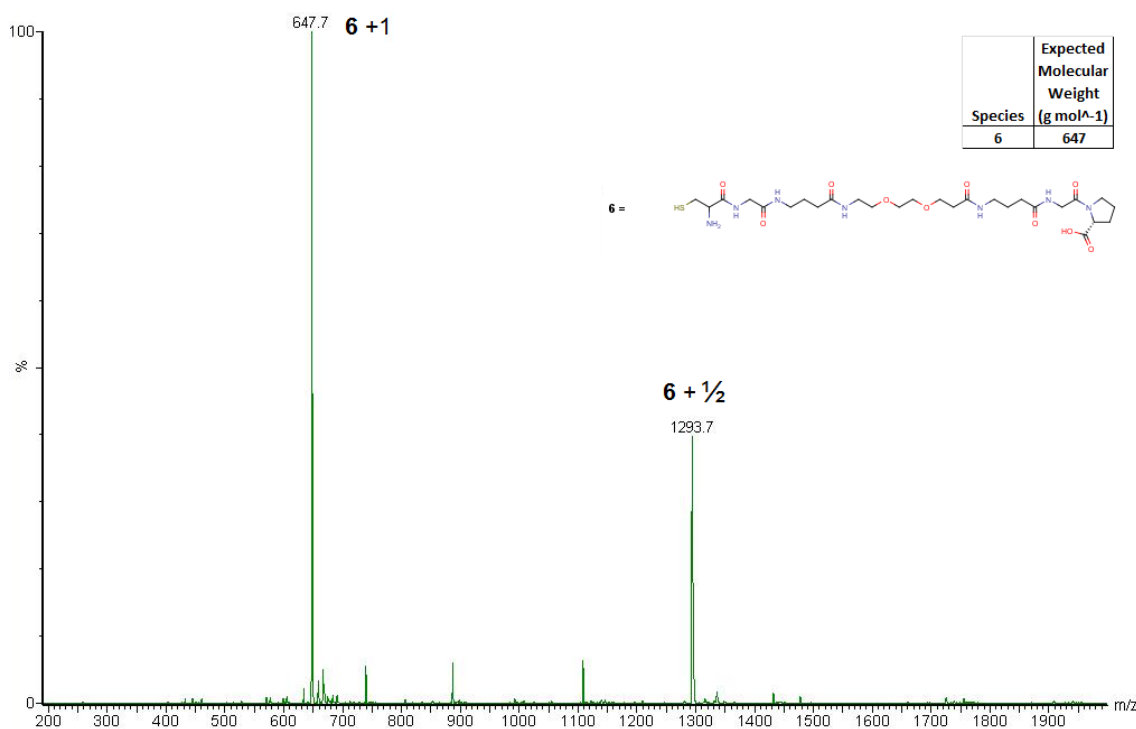
#### 4.3.3. Central scaffold cyclisation by thioether formation

Disulphide bond formation was expected to be an issue in the formation of thioether bonds, but it was thought that the dimerisation reaction would not be significant in the case of the intramolecular cyclisation of **3** if the reactants were kept sufficiently dilute. This proved to be the case with no disulphide bond formation detectable by MS. The

mass spectrum of the cyclised material showed peaks for a  $M+3H$  ion at 514  $m/z$  and an  $M+2H$  ion at 771  $m/z$  (see appendix), which corresponds to a loss of 36 mass units from the uncyclised material (peaks:  $M+3H$  ion: 526  $m/z$ ,  $M+2H$  ion: 789  $m/z$ ) which is the mass change expected from the loss of HCl upon intramolecular thiol alkylation by an alkyl halide. Furthermore, an Ellman test of the purified material showed no colour change. These two experiments provide confidence that the desired reaction was successfully performed and product **4** had the correct structure post-cyclisation.

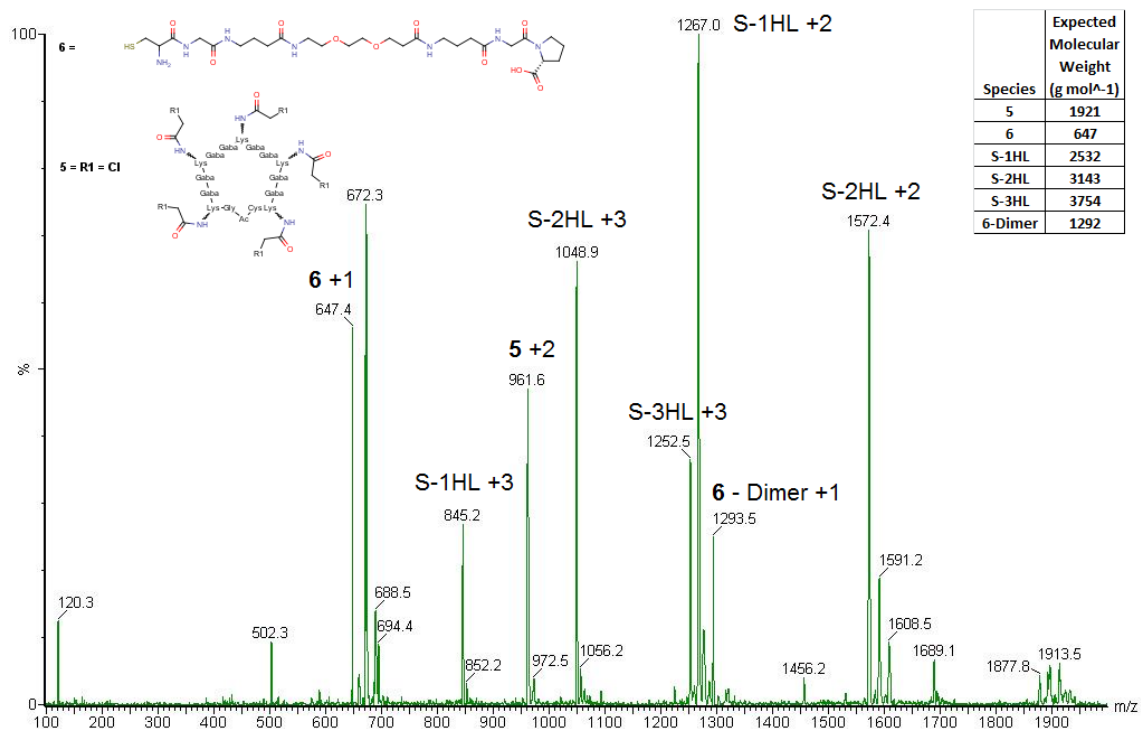
#### 4.3.4. Multivalent ligand synthesis by thioether formation

The initial synthesis of **21** was attempted at room temperature open to the atmosphere to investigate if significant dimerisation of **6** did occur in air as expected. A mass spectrum was recorded at 2 hours after the reaction had started (Figure 4.18) with the original mass spectrum of **6** shown in Figure 4.17.

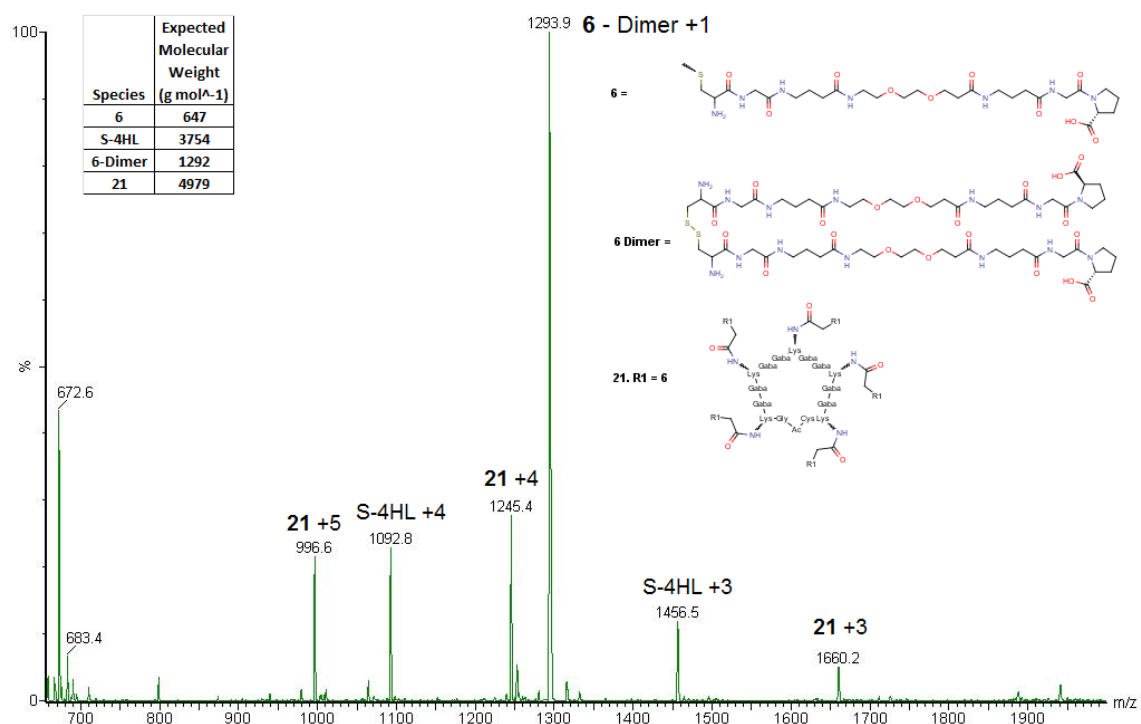


**Figure 4.17.** Mass Spectrum of purified linker-headgroup **6** prior to reaction with central scaffold **5**.





**Figure 4.18.** Mass Spectrum of reaction between **5** and **6** in air after 2 hours. Various partially reacted central scaffold units are visible with as many as 3 linker-headgroups attached. Some dimerisation of the linker-headgroups is visible, as is starting material **5**.



**Figure 4.19.** Mass spectrum of reaction of **5** with **6** in air after 7 hours. The most abundant peak corresponds to dimerisation of linker-headgroup compound group **6** by disulphide formation. This was due to performing the reaction in air. The desired product **21** is visible and the multivalent product

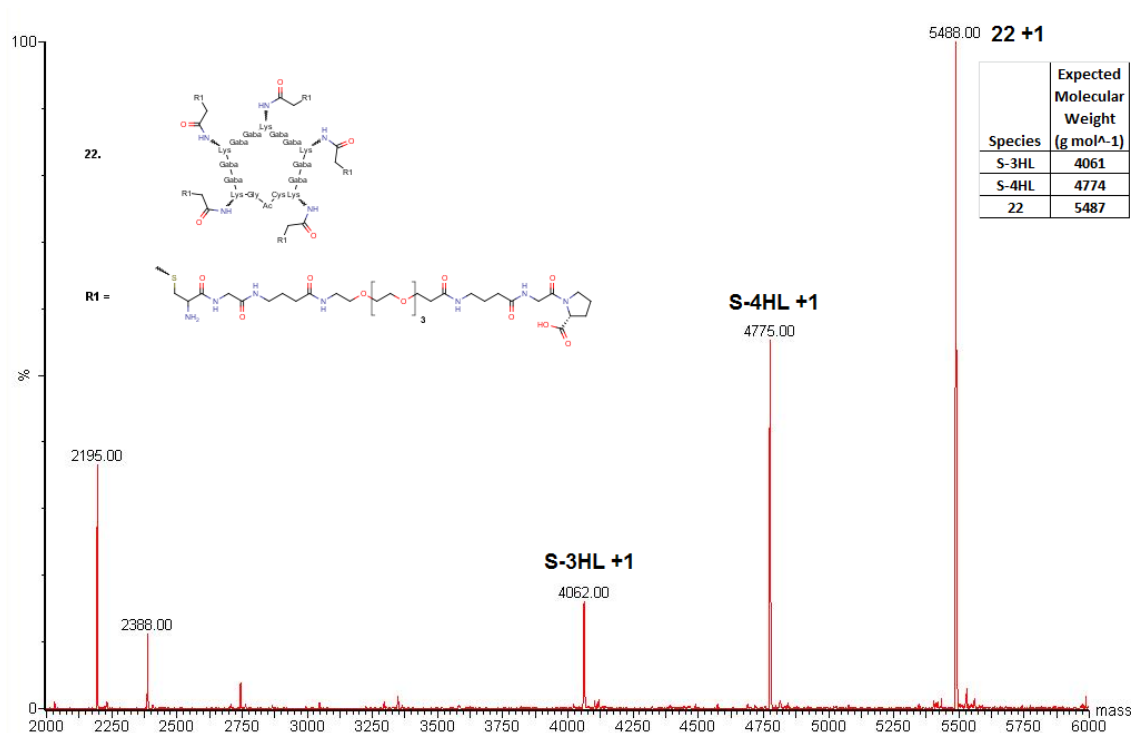
missing a single LH, no dehalogenation is seen from the unreacted site on the central scaffold.

After 2 hours, the original peak for the monomeric ligand at 648.6 m/z had been replaced by another at 647.4 and a second peak corresponding to the mass of two headgroup molecules appeared at 1293.5 m/z indicating that significant dimerisation had occurred. However, there were clear signs of the desired reaction with S-1HL, S-2HL and S-3HL all detectable by MS. It was also clear that no de-halogenation had occurred to the remaining chloroacetyl groups on the partially reacted multivalent ligands. Therefore, further **6** was added and the reaction left for a further 5 hours before another MS, which subsequently showed the desired product to be the principal multivalent species present (Figure 4.19). However, a very large peak on the MS for the dimerised product was seen and the spectrum had to be cropped at 650 m/z because the 647 peak of the half charged ion of dimerised LH vastly exceeded the relative abundance of the detected multivalent product, making it difficult to assess the extent of the desired reaction.

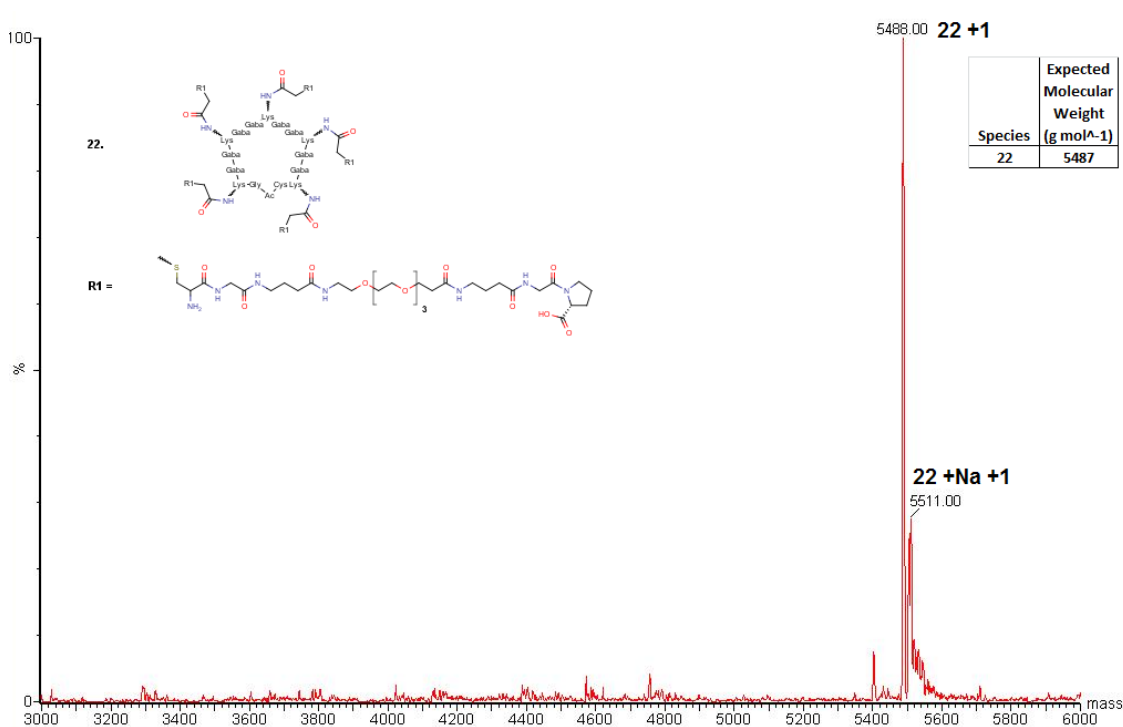
The product was then taken, freeze-dried and purified by HPLC. Anecdotally, the MS of the product after drying appeared to show greater progression of the reaction to the desired S-5HL product, suggesting that further thiol alkylation had occurred on the freeze-dryer. HPLC purification showed a very large peak for the dimerised **6** some remaining monomeric **6** and a range of peaks for S-nHL, of which the peak for S-5HL was the largest. These results gave confidence that the reaction was possible and would proceed without difficulty if ligand-headgroup dimerisation could be minimised. The anecdotal observations of the reaction occurring on the freeze-dryer provided an innovative solution to the dimerisation problem.

A new reaction to synthesise product **22** was prepared by mixing **5** and linker-headgroup **7** as before, but immediately after addition of base, the reaction was frozen at -20°C and placed in a dome freeze dryer at -44°C under vacuum. After drying, the sample was analysed by MS and showed conversion the majority of **5** to final multivalent product with some S-4HL and S6-3HL (Figure 4.20). A small sample was tested for free thiol groups using Ellman's reagent, which produced a yellow colour indicating not all the headgroup had dimerised. Therefore, the sample was immediately redissolved in 10ml 1:1 acetonitrile:water, the pH adjusted to 8.0 with sodium

bicarbonate and freeze-dried once more. The resulting mass spectrum of the second freeze-drying showed only the desired multivalent product (Figure 4.21). On this basis all other multivalent ligands were synthesised by repeated freeze-drying until no deficiency in the number of headgroups was detected.



**Figure 4.20.** Maximum entropy mass spectrum of reaction between **5** and **7** in air after drying on freeze-dryer.



**Figure 4.21.** Maximum entropy mass spectrum of reaction between partially formed **22** and **7** after second freeze-drying of reactants.

#### **4.4. Discussion**

The main aim of this project was to synthesise multivalent ligands for SAP and although it was decided that peptides should constitute a major part of the structure, there were no additional constraints over the nature of the chemical linkages used to form the multivalent constructs. This permitted the use of thioether bonds, which have been shown to be applicable in both intramolecular cyclisation and intermolecular ligation reactions for constructing multivalent scaffolds.

##### **4.4.1. Forming thioether bonds during freeze-drying**

As far as this author is aware, this document constitutes the first case of thioether formation during freeze-drying the reaction mixture. This method provides the benefit of combining the reaction and sample desolvation steps. Furthermore, the concentration of the reactants actually increases as the reaction proceeds due to the shrinking volume of solvent, in contrast to conventional solution phase reactions where the concentration of reactants falls as they react in a constant volume of solvent. Naturally, the converse of this is that the reactants are expected to have greatly reduced freedom of movement in the solid phase. It is unclear whether the reaction is truly occurring in the solid phase or whether the sample reaches a stage where the oily nature of the reactants acts causes them to act as their own solvent when approaching dryness.

A drawback of the freeze-drying method is that the entire process of solvation, freezing and drying must be repeated if the reaction is not complete and the reaction cannot be monitored while in process. Greater optimisation of the reaction conditions may improve the reaction efficiency and make the process quicker than performing the thioether reaction and freeze-drying steps separately. However, the maximum benefit of these reaction conditions may be attained by using them in conjunction with a normal room temperature reaction under nitrogen. Reactions do not pay attention to the working hours of the laboratory scientist and often require overnight to reach completeness. The solvent must then be removed the following day, which means the whole process takes much longer to complete. Using the reaction methodology presented here, a faster hybrid approach can be employed where the thioether reaction is started at room temperature and then continued overnight on a freeze-dryer yielding the

desired product in solid form ready for purification. Therefore, with further investigation it is possible that this method will add greater speed and flexibility in the synthesis of thioether containing compounds.

#### 4.4.2. Identification of multivalent ligands by mass spectrometry

Mass Spectrometry (MS) is the standard method of sample identification in peptide synthesis. Electrospray Ionisation Mass Spectrometry (ESI-MS) was the principal technique used to characterise the multivalent ligands produced in this investigation, with additional qualitative analysis using Ninhydrin and Ellman's reagent to identify the presence of free primary amines and free thiol groups, respectively. Soft ionisation techniques, such as ESI, charge a peptide while minimising fragmentation, making it possible for the mass of the entire molecule to be analysed accurately.

The final mass spectra of the purified multivalent ligands shows peaks corresponding to molecules carrying multiple charges, which is to be expected for relatively large peptides. Maximum entropy calculations can deconvolute multiply-charged peaks to show the molecular weight of the species that produced recorded spectrum (Ferrige *et al.*, 1991). Maximum Entropy calculations were performed using the Micro Mass software with the results for each multivalent ligand detailed in Table 4.3 next to the predicted mass. The observed masses are in line with the predicted masses, with small variations caused by protonation of the sample in the MS. Furthermore, the mass difference between the individual components and the combined product masses are exactly as expected. The addition of five LH to the compound **5** resulted in a mass that is  $MW_5 + 5(MW_{LH}) - 5(MW_{HCl})$  where  $MW_{HCl}$  is the molecular weight of HCl formed during the thiol alkylation. This calculation is evidence that the desired thioether bond has formed and the correct product has been synthesised. Further confidence of this conclusion is provided by the results of the Ellman's tests performed on the LH components prior to reaction with **5** and another test performed on the final products.

Ligand	Expected Mass(g mol <sup>-1</sup> )	Expected Mass			Observed Mass(g mol <sup>-1</sup> )	Observed Mass		
		M+4H	M+5H	M+6H		M+4H	M+5H	M+6H
<b>21</b>	4973	1244	996	830	4980	1245	997	–
<b>22</b>	5488	1373	1098	916	5492	1373	1099	916
<b>23</b>	5928	1483	1186	989	5932	1483	1187	989
<b>24</b>	6178	1545	1236	1031	6184	1546	1237	1031
<b>25</b>	6693	1674	1339	1116	6695	1674	1339	1116
<b>26</b>	7133	1783	1427	1188	7136	1784	1427	1189

**Table 4.3.** Predicted and observed masses for multivalent ligands **21-26**. Predicted figures were calculated based on the molecular weight of the constituent residues. The actual masses observed for final purified products were measured by maximum entropy mass spectrometry. The masses of the observed products are slightly higher than the calculated mass due to protonation during sample preparation. These variations are not considered significant.

On the basis of these results, there is little doubt that multivalent ligands of the desired structure have been produced with sufficient quantity and quality for further investigation into their potential for binding with SAP. The results also demonstrate the versatility of the thioether linkage in the synthesis of multivalent ligands. In this investigation they were used for both intra- and intermolecular bond formation using standard, inexpensive and benign chemicals in reactions that proceeded with minimal side reactions when appropriate reaction conditions were applied.

#### 4.4.3. Avoiding diketopiperazine formation

Diketopiperazine (DKP) formation has been shown to be an issue for the synthesis of pentavalent ligands **21**, **22** and **23**. The D-proline at the C-terminus will exacerbate DKP formation like no other standard amino acid. It was thought that the use of 2-chlorotrityl chloride resin would remedy the issue, but clearly the presence of 5 linkers on each central scaffold has magnified the influence of whatever small quantity of DKP had occurred. The issue can be easily overcome in future by simply placing GABA as the second amino acid rather than Gly and then placing Gly as the third amino acid. DKP requires the formation of a six-member heterocycle from the three backbone atoms of two amino acid residues, which requires two  $\alpha$ -amino acids at the C-terminus.

## **Chapter 5. Analysis of binding between SAP and multivalent ligands 21-26**

### **5.1. Introduction**

The multivalent ligands synthesised in Chapter 4 were analysed for calcium-dependent binding to SAP by experiments designed to answer a logical series of questions regarding their interaction with the protein. Firstly, Ultraviolet-Visible spectroscopy was employed to determine whether the multivalent ligands bound at all to SAP. Secondly, mass spectrometry and x-ray crystallography were used to determine the mode of any observed binding, in particular whether the ligands bind through their carboxyl termini as seen for univalent D-proline in Chapter 2 and evaluate whether the decavalent ligands preserve the ability to cross-link two SAP pentamers in the non-covalent manner seen with CPHPC (Pepys *et al.*, 2002). The final step, having determined if and how binding occurs, was to calculate the strength of any binding interaction using isothermal titration calorimetry (ITC).

ITC requires accurate determination of sample quantities and purity (Wisemann *et al.*, 1989). Although the purity of the multivalent peptide ligands was calculated by LCMS in Chapter 4, the percentage peptide content of the sample remained undetermined. Amino acid analysis was selected as the appropriate technique for calculating peptide content because the multivalent ligands lack functional groups active in the UV/Vis spectrum necessary for standard absorbance measurements.

### **5.2. Materials and Methods**

#### **5.2.1. Peptide content analysis**

Purified, solid peptides still contain a considerable quantity of counter ions, water of crystallisation and other non-peptide impurities. Accurate measurements of peptide content are particularly important for ITC where assuming the weighed mass of a sample is 100% peptide will almost certainly lead to a gross underestimation of binding affinity. Protein concentrations are typically determined by measuring absorbance of UV light by aromatic side chains of Tyr and Trp residues at 280nm in a UV/Vis



spectrometer. The multivalent ligands do not contain aromatic groups, so an alternative strategy was required. In this investigation, amino acid analysis was employed to measure the total peptide content of purified multivalent ligand samples and LCMS data from Chapter 4 was used to measure the proportion of the desired product in the total peptide content.

Amino acid analysis is typically employed for analysing amino acid contents of unknown polypeptides or for peptide sequencing by Edman degradation (Edman *et al.*, 1950). Amino acid analysis uses concentrated (6M) HCl heated at 100°C for 20 hours to first breakdown peptides into their individual amino acids. The amino acid mixture is separated by ion exchange chromatography and the eluate reacted with ninhydrin to facilitate peak detection. The elution time for each peak is compared to retention times for known amino acid standards to identify the amino acid constituents of the sample. It is possible to quantify the exact peptide content by comparing the observed peak area with the expected values based on the quantity of sample used (Spackman *et al.*, 1958).

Amino acid analysis was performed by Peter Sharratt at the Protein and Nucleic Acid Chemistry Facility (University of Cambridge, UK). In a typical experiment, 0.5-1.0mg accurately weighed peptide is dissolved in 1ml 10% acetic acid and mixed thoroughly. A norleucine internal standard is added to a 20µl aliquot of the solution and the mixture transferred to a pyrolysed tube. The tube is placed in a centrifugal evaporator and the mixture concentrated to dryness. Constant boiling hydrolysis is performed with gas phase hydrochloric acid in a hydrolysis vial at 115°C for 22 hours. After removing traces of acid, sodium citrate loading buffer, pH 2.2, is added to dissolve the residue. The resulting solution is filtered under centrifugation through a 0.2 micron filter. An aliquot of the filtrate is injected onto a Biochrom 30 amino acid analyser and chromatography performed on an ion exchange resin, utilising a sodium system, eluting with a series of buffers over the pH range 3.2 to 6.5. Peak detection is achieved by mixing the eluate with ninhydrin at 135°C and measuring the absorbance at 570 and 440nm. Quantitation is performed using Chromeleon software and calibration curves for each amino acid of interest (correspondence with Peter Sharratt, University of Cambridge, UK; Spackman *et al.*, 1958). In the analysis of the ligands **21-26**, the

observed values of Gly, GABA, Lys and Pro were used to determine the peptide content of the samples.

### **5.2.2. SAP ligand precipitation assay using Ultraviolet-Visible Spectroscopy**

As mentioned previously, SAP auto-aggregates in the presence of calcium unless there is an SAP binding ligand present (Pepys *et al.*, 1997). This property can be exploited to rapidly screen ligands for binding to SAP. Typically, this is done by mixing microlitre drops of SAP stock solution with an excess of ligand at physiological pH, adding a small volume of 2mM calcium chloride solution and observing the drop under a microscope to look for precipitation, which denotes a lack of binding. Performing the analysis with larger volumes and looking for absorption of light by samples in a UV/Vis spectrometer provides a more precise method for analysing binding. Ligands are mixed with SAP and incubated in calcium-containing or calcium-free conditions and their absorption compared to a ligand-free, calcium-containing sample of SAP.

A 2.0 $\mu$ M multivalent ligand solution was buffered to pH 8.0 with 90mM Tris. 5 $\mu$ l aliquots of the multivalent ligand solution were mixed with 5 $\mu$ l of 10.1 mg ml<sup>-1</sup> (0.4 $\mu$ M) SAP in TN buffer (10mM Tris, pH 8.0, 140mM NaCl, 0.1% w/v NaN<sub>3</sub>) and 5 $\mu$ l of either 600mM calcium chloride (calcium-containing assay) or 5 $\mu$ l TE buffer (10mM Tris, pH 8.0, 140mM NaCl, 10mM EDTA, 0.1% w/v NaN<sub>3</sub>; calcium-free assay). After incubating for 3 hours at room temperature, the calcium-containing samples were diluted with TC buffer (2mM CaCl<sub>2</sub>, 10mM Tris pH 8.0, 140mM NaCl) and the calcium-free samples diluted with TE buffer to a final concentration of 1ml. The samples were mixed in a vortex for 1 minute and transferred to cuvettes.

Absorption of 320nm wavelength UV light by samples was measured 5 minutes after dilution and then again 35 minutes after dilution. Measured absorption values were compared to ligand free samples containing 5 $\mu$ l 10.1mg ml<sup>-1</sup> SAP in TN buffer and 995 $\mu$ l TC buffer.

### 5.2.3. Mass spectrometry of SAP-multivalent ligand mixture

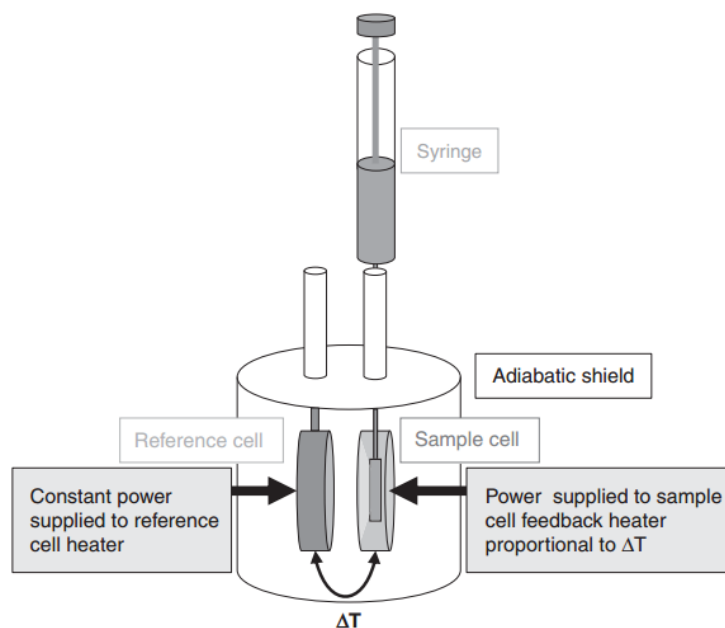
Mass spectrometry (MS) of whole SAP-decavalent ligand complexes was performed by Idir Liko at the University of Oxford, UK. Whole protein MS has a variety of uses in proteomics including analysing protein-ligand interactions (Aebersold and Mann, 2003). The key aspects of calibration of equipment, sample preparation and analysis of results have been detailed extensively by Henández and Robinson (2007).

#### 5.2.3.1. MS sample preparation

Samples were prepared by diluting  $10.1\text{mg ml}^{-1}$  ( $81\mu\text{M}$ ) SAP in  $200\text{mM}$  ammonium acetate, pH 7.4 to a final concentration of  $0.9\mu\text{M}$ . A two-fold excess of multivalent ligand ( $1.8\mu\text{M}$ ) was added to the protein solution and incubated at room temperature. After 1 hour,  $5\text{mM}$  calcium chloride was added and the samples biospun in  $200\text{mM}$  ammonium acetate, pH 7.4, ready for MS analysis. MS was performed on a Waters QToF Mass Spectrometer with a  $1.5\text{kV}$  capillary voltage,  $60\text{V}$  cone voltage and a collision (SID) of  $20\text{V}$ .

### 5.2.4. Isothermal titration calorimetry of SAP with multivalent ligands

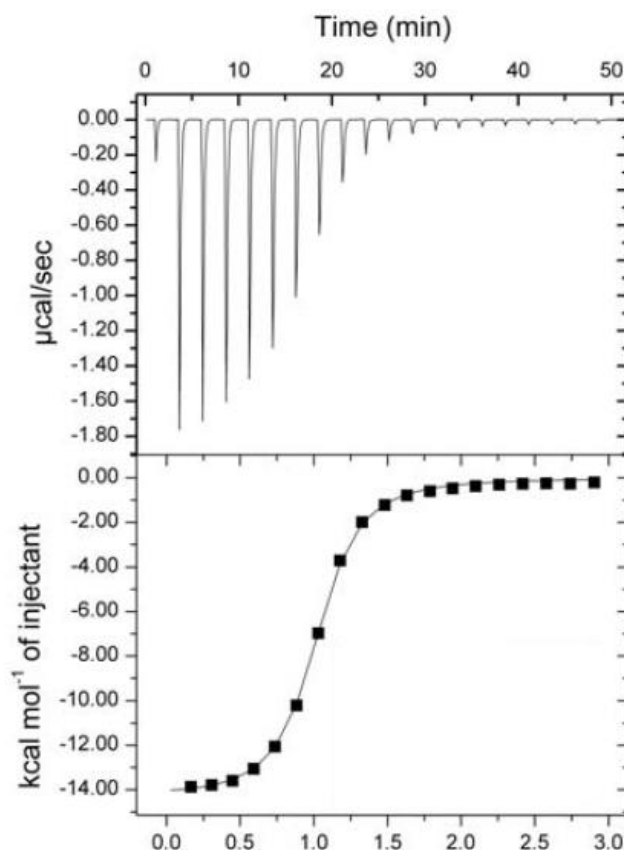
Isothermal titration calorimetry (ITC) is a powerful technique that can be used to measure binding association constant  $K_a$ , enthalpy,  $\Delta H$ , and stoichiometry,  $n$ , of ligand-protein interactions (Wisemann *et al.*, 1989). The calorimeter measures the heat change during a binding event and consists of two cells, a sample cell and a reference cell, made from a highly conductive material such as Hastelloy or gold (Pierce *et al.*, 1999). The two cells are encased within an adiabatic jacket, preventing heat transfer with the outside environment. Heaters on both cells are used to raise the temperature to a specific value, typically  $25^\circ\text{C}$ , and the power applied to maintain the temperature is recorded as the reference baseline signal. The protein of interest is added to the sample cell in a buffer at a precise concentration prior to the start of the experiment. The ligand, prepared in identical buffer conditions is added periodically throughout the experiment using a syringe specially designed to rotate with a paddle-shaped needle to mix the cell constituents (Figure 5.1).



**Figure 5.1.** Typical design of calorimeter used in isothermal titration calorimetry. Adapted from Freyer and Lewis, 2008).

As the protein is titrated with the ligand, a heat change occurs depending on whether the reaction is endo- or exothermic. Using an exothermic reaction as an example, the heat evolved increases the temperature of the sample cell relative to the reference cell. The calorimeter detects this temperature mismatch with thermopile/thermocouple circuits and reduces the power applied to the heating element on the sample cell until the two cells are returned to the same temperature (Freyer and Lewis, 2008). ITC records the time-dependent input of power required to maintain equal temperatures between the cells during a titration in units of  $\mu\text{cal s}^{-1}$ .

Plotting the  $\mu\text{cal s}^{-1}$  measurements for each injection gives a graph with peaks whose magnitude decreases during the titration as the binding sites of the protein become saturated with ligand (Figure 5.2). Integration of the area of each peak gives the total heat change of each injection, which can be plotted against the molar ratio of  $[L]_{\text{total}}/[M]_{\text{total}}$ , where  $[L]_{\text{total}}$  and  $[M]_{\text{total}}$  are the total concentration of ligand and free macromolecule in the sample cell. Fitting a sigmoidal curve to the data points by non-linear least squares calculation gives a curve from which  $K_a$  and  $\Delta H$  can be obtained and where the value on the x-axis at the midpoint of the curve is the stoichiometry,  $n$ , of the binding reaction (Figure 5.2).



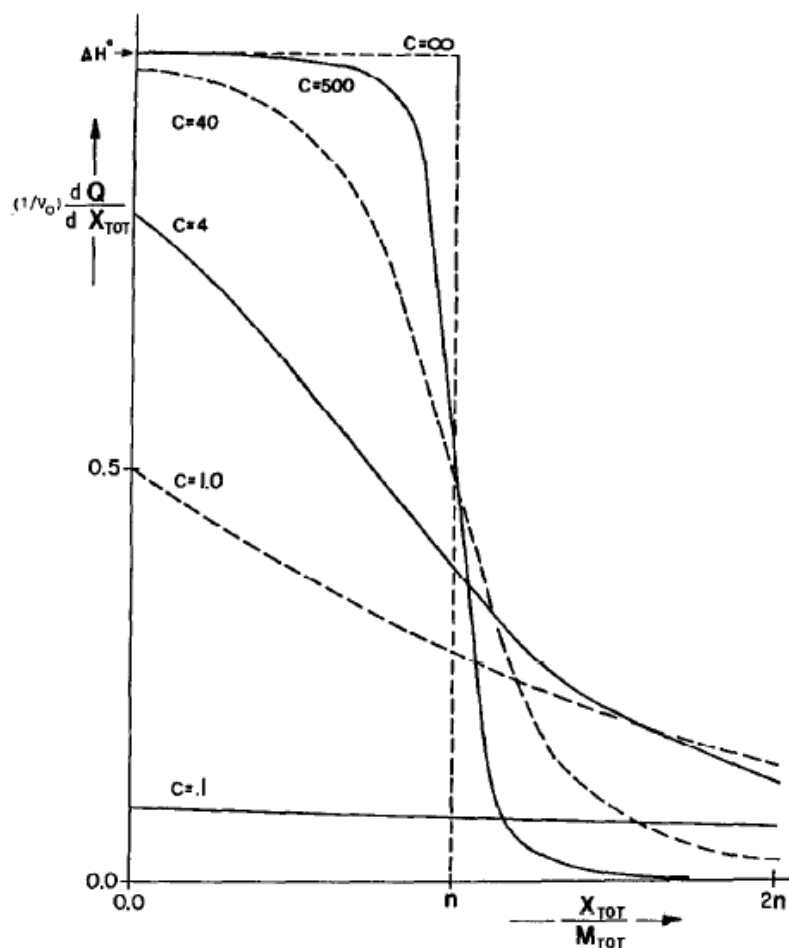
**Figure 5.2.** Typical binding isotherm obtained from an ITC binding experiment. Top: raw heat change data recorded for each injection. Bottom: plot of heat evolved/taken in against the molar ratio of ligand to protein with a red line showing the sigmoidal line of best fit calculated by non-linear least squares (MicroCal ITC 200 User Manual, GE Life Sciences, Uppsala, Sweden, 2012).

The shape of the curve in an ITC experiment is determined by the concentration of the macromolecule,  $[M]$ ,  $K_a$  and  $n$ . Modern ITC software can predict the curve recorded for a binding experiment if realistic estimates can be made of  $K_a$  and  $n$ , this can be used to optimise the protein concentration based on the expected binding affinity. The influence of  $[M]$ ,  $K_a$  and  $n$  on the expected binding curve can be expressed using the unitless  $c$  value in the relationship:

$$c = K_a \cdot [M] \cdot n \quad \text{Equation 5.1.}$$

$c$  tends towards infinity as strength of binding increases, producing a curve that has an

instantaneous transition from unsaturated to complete saturation of all binding sites within a single injection as shown by Figure 5.3. This leads to a lack of data at the transition point making accurate  $K_a$  determination impossible, although  $\Delta H$  and  $n$  can still be calculated accurately. Figure 5.3 shows the expected binding curve for a range of  $c$  values (Wisemann *et al.*, 1989). Weak binding also prevents accurate determination because the curve's sigmoidal shape is lost. The most accurate ITC data is obtained when the  $c$  value is between 1-1000, this corresponds to  $K_a$  values in the range of  $10^3$ - $10^9$   $M^{-1}$  (O'Brien *et al.*, 2000). Larger  $K_a$  values would require the protein to be diluted to a point where the heat change upon addition of the ligand would be less than the detection limit of the apparatus (typically  $0.1 \mu cal\ s^{-1}$ ; Freyer and Lewis, 2008).



**Figure 5.3.** Influence of  $c$  value (Eq. 5.1) on shape of sigmoidal curve (adapted from Wisemann *et al.*, 1989). As  $c$  value increases, the number of data points obtained on a curve decreases, making accurate determination of the binding constant from the gradient of the slope impossible. However, the enthalpy calculation becomes more accurate, as this is calculated from the heat change from the initial injections; as long as complete saturation of the binding sites does not occur in initial phase of the experiment.

From this point on binding affinities will be referred to in terms of the dissociation constant,  $K_d$ , which is the reciprocal of  $K_a$ . The reasons for this are: firstly, that all previous SAP ITC experiments report binding affinities in terms of  $K_d$ , following this convention will provide greater clarity of the results. Secondly, it can be shown mathematically that when half the binding sites of the protein are occupied, the dissociation constant is equal to  $[L]$ , the concentration of unbound ligand in the sample cell (derivation found in appendix section A.4); this makes the  $K_d$  a convenient way to visualise how the binding constant actually relates to the reaction occurring in the sample. Simply put, the smaller the dissociation constant, the lower the concentration of unbound ligand at half-occupancy.

#### 5.2.4.1. ITC sample preparation

Sample preparation for ITC requires precise buffer matching between the ligand and protein samples. The measured heat change in a binding reaction can be expressed as the following:

$$\Delta H = \Delta H_{bind} + \Delta H_{conf} + \Delta H_{ion} \quad \text{Equation 5.2.}$$

The observed enthalpy change ( $\Delta H$ ) is the sum of enthalpy contributions from binding,  $\Delta H_{bind}$ , conformation changes,  $\Delta H_{conf}$ , and ionisation  $\Delta H_{ion}$ . Buffer mismatch introduces a systematic error by increasing the heat contribution from ionisation. This is a particular issue for Tris-Cl buffer, which has a relatively large heat of ionisation (O'Brien *et al.*, 2001). Furthermore, during a binding assay, part of the heat change contribution is from by-products of the binding reactions, namely from dilution of the ligand upon injection into the sample cell and other non-specific effects (Blandamer, 1998). These effects were removed by first performing a blank experiment to measure the background heat change from injecting each ligand into buffer and then using this data to subtract from the experimental data during data analysis. Equation 5.3 demonstrates the calculation performed in the experimental data:

$$\Delta q_i = \Delta q_{app} - \Delta q_{dil} - \Delta q_{ns} = n[M]_{tot} \cdot V_{cell} \cdot \Delta H_{app} \cdot R \quad \text{Equation 5.3.}$$

Subtracting background measurements of heat contributions from ligand dilution,  $\Delta q_{dil}$ , and non-specific effects,  $\Delta q_{ns}$ , from raw heat change data for injection  $i$ . The heat change for the  $i$ th injection is equal to the macromolecule concentration,  $[M]$ , which has  $n$  identical independent binding sites, in total cell volume,  $V_{cell}$ , where the enthalpy of binding is  $\Delta H_{app}$  and  $R$  represents the degree of saturation of the macromolecule binding sites.

Precise knowledge of reactant concentrations is also key for calculating accurate results from raw data in ITC. As previously mentioned, amino acid analysis was undertaken to establish exact peptide content of the multivalent ligand samples (Section 5.3.1), while the concentration of SAP stock solutions were measured using UV/Vis as detailed in the following section.

#### **5.2.4.2. ITC of SAP with multivalent ligands 22-26**

Auto-aggregation of SAP in the presence of physiological calcium concentration prevents ITC measurements under physiological conditions (Pepys *et al.*, 1997). However, Kolstoe found that precipitation can be prevented using a concentration of calcium ions in excess of 200mM, most likely due to the high concentration of chloride ions coordinating calcium ions bound to SAP preventing binding by Glu167 by other SAP pentamers (Kolstoe thesis, 2005). Both SAP and the multivalent ligands were prepared in high calcium TC-like buffer containing 200mM  $\text{CaCl}_2$ , 140mM NaCl, 10mM Tris, pH 8.0, 0.01% w/v  $\text{NaN}_3$ . SAP samples were prepared at a concentration of 0.0096mM, the exact protein concentration of the initial stock solution was determined using by UV/Vis measurement at 280nm using an extinction coefficient of  $1.71\text{M}^{-1}\text{cm}^{-1}$  as determined by de Beers and Pepys (1982). The multivalent ligand samples were prepared for injection at a concentration of 0.080mM for pentavalent ligands **22** and **23**, and between 0.032-0.040mM for decavalent ligands **24-26**. The reference cell was filled with water and 0.01% w/v  $\text{NaN}_3$ . All samples were degassed and spun down in a centrifuge prior to injection.

ITC experiments were performed using a GE MicroCal 200 at 25°C with a reference



power of  $6\mu\text{cal s}^{-1}$  and a mixing of 1000 rpm. 280 $\mu\text{l}$  water, 0.01% w/v  $\text{NaN}_3$  was injected into the reference cell and the same volume of SAP in high calcium TC buffer injected into the sample cell. The titration syringe was filled with 40 $\mu\text{l}$  multivalent ligand solution in high TC buffer. An initial 0.4 $\mu\text{l}$  dummy injection was performed to counteract diffusion of ligand from the injector tip during equilibration. Subsequently, the ligand was injected into the sample cell in 1 $\mu\text{l}$  volumes for the pentavalent ligands and 2 $\mu\text{l}$  for decavalent ligands at a rate of 2 times the injection volume per second, injections were separated by 120 seconds. Analysis of results was conducted with Origin software using a single binding site model (version 7.20, MicroCal, GE Life Sciences, Uppsala, Sweden).

#### **5.2.4.3. ITC of SAP with **21****

Multivalent ligand **21** was analysed by ITC as part of a proof of concept study prior to synthesis of **22-26**. ITC of this ligand was performed on a TA instruments NanoITC. The preparation of samples for **21** was the same as for those analysed on the MicroCal apparatus, but the protein and ligand quantities were adjusted due to lower sensitivity of the calorimeter and the larger sample cell of the NanoITC. 1ml 0.008mM SAP in high calcium TC buffer was added to the sample cell, into which 0.16mM ligand **21** in identical buffer was injected using 8 $\mu\text{l}$  injection volumes separated by 200s after an initial 4 $\mu\text{l}$  injection. The results were analysed using NanoAnalyze (version 2.2.3, TA instruments, New Castle, Delaware).

#### **5.2.5. X-ray crystallography of SAP with multivalent ligands **23** and **24****

##### **5.2.5.1. Crystallisation**

Separate crystal screens were prepared to co-crystallise each of the five multivalent ligands with SAP by hanging-drop vapour diffusion at room temperature, but SAP crystals were only obtained in screens containing ligands **23** and **24**. Crystals were cryoprotected by sequential 5 minute immersions in microlitre drops of crystallisation cocktail containing steadily increasing concentrations of glycerol, up to a final concentration of 20% v/v glycerol (Garman, 1999). Crystals were then harvested in a nylon loop, flash cooled in liquid nitrogen and stored at 100K prior to data collection.

#### **5.2.5.2. Data collection**

Data sets for crystals of both multivalent ligands in complex with SAP were collected at 100K on beamline I03 at Diamond Light Source (Didcot, UK). **SAP-23**: 1500 images were collected with an oscillation angle of  $0.2^{\circ}$  from a single crystal using 50% transmission of the x-ray beam and an exposure time of 0.04s. **SAP-24**: 2400 images were collected with a  $0.15^{\circ}$  oscillation angle from a single crystal using 20% transmission of the x-ray beam with an exposure time of 0.05s.

#### **5.2.5.3. Data processing, molecular replacement and structure refinement**

Data for the **SAP-23** crystal was processed using XDS and XScale as part of the Xia2 automatic data processing 3daii run (Kabsch, 2010), merged with Aimless (Evans, 2006), scaled using Scala. For **SAP-24** data processing was completed using Mosfilm (Powell, 1999). Molecular replacement for both datasets was completed using Phaser with PDB file 1SAC used as the search model (McCoy *et al.*, 2007; [www.pdb.org/rcsb](http://www.pdb.org/rcsb)), followed by a single round of rigid body refinement in Refmac (Murshudov *et al.* 1997). Model building was undertaken in Coot (Emsley and Cowtan, 2004). Subsequent rounds of restrained refinement for **SAP-24** were completed using Refmac applying global NCS restraints. Coordinate files for fractions of ligand **24** were generated using the Prodrgr server (Schüttelkopf and Van Aalten, 2004) with geometric restraints calculated with Phenix.elbow (Afonine *et al.* 2012).

### **5.3. Results**

#### **5.3.1. Peptide content analysis**

Peptide content analysis was performed on all multivalent ligands to give accurate quantities for use in subsequent experiments, in particular isothermal titration calorimetry, where exact quantities are needed to determine meaningful binding constants and reaction stoichiometry. The results of the amino acid analysis are summarised in Table 5.1, the average peptide content ranged between 46.0% for **21** and 78.5% for **22**. Peptide content analysis was not performed prior to ITC of ligand **21** as it was synthesised as part of a preliminary proof of concept study preceding synthesis of

ligands **22-26**.

<b>Multivalent Ligand</b>	<b>Peptide Content (%)</b>		
	<b>Run 1</b>	<b>Run 2</b>	<b>Average</b>
<b>21</b>	49.0	43.0	46.0
<b>22</b>	77.0	78.0	77.5
<b>23</b>	61.0	62.0	61.5
<b>24</b>	57.0	58.0	57.5
<b>25</b>	68.0	65.0	66.5
<b>26</b>	68.0	64.0	66.0

**Table 5.1.** Peptide content of multivalent ligand samples as determined by amino acid analysis.**5.3.2. UV-Vis spectroscopy SAP precipitation assay**

The auto-aggregation of SAP is caused by the side chain carboxyl group of Glu167 on the A-face of each subunit coordinating the double calcium binding site on the B-face of another pentamer (Emsley *et al.*, 1994). Aggregation is completely inhibited by the presence of a SAP binding ligand such as N-acetyl D-proline (Pepys *et al.*, 1997). Using UV/Vis spectrometry to measure absorption, this property can be exploited to rapidly screen compounds for calcium-dependent binding to SAP. If the potential ligand does not bind to SAP, precipitation will occur leading to a rise in absorption, otherwise the absorption should remain at the baseline level determined from samples containing SAP and ligand, but calcium-free. The measured absorption value is also compared to an aggregated SAP sample (ligand-free SAP in TC buffer).

Around 50µg SAP is incubated with a 5-fold excess of multivalent ligand either in the presence or absence of calcium. During incubation 200mM CaCl<sub>2</sub> is used to because the high concentration of calcium prevents SAP precipitation. The incubated samples are then diluted to give a final 1ml volume and physiological calcium concentration.

Tables 5.2 and 5.3 show that no significant precipitation was observed for any of the calcium containing solutions indicating that all the multivalent ligands bind to SAP with sufficient affinity to prevent auto-aggregation. A small increase in absorption was observed for **26** indicating that some precipitation had occurred. However, Run 1 of the

calcium free sample for **26** showed a similar rise above the baseline level suggesting that the increase may not have been due to SAP precipitation. Furthermore, the signal is still some 8-fold smaller than the absorption of auto-aggregated SAP in the ligand free, TC buffer samples. These results were taken as evidence of binding and that further analysis of their binding to SAP was warranted.

Ligand	Buffer	Absorption at 320nm			Average
		Run 1	Run 2	Run 3	
Water Blank	None	0.000	0.000	0.000	0.000
Ligand Free SAP	TC	0.137	0.120	0.192	0.150
S6-5H3	TC	-0.002	-0.003	0.000	-0.002
	TE	0.000	-0.001	0.000	0.000
S6-5H5	TC	0.003	0.001	0.003	0.002
	TE	-0.003	0.000	0.000	-0.001
S6-5H110	TC	0.003	0.002	0.007	0.004
	TE	-0.003	0.000	0.000	-0.001
S6-5H310	TC	0.001	-0.002	-0.001	-0.001
	TE	0.003	0.003	-0.003	0.001
S6-5H510	TC	0.001	0.015	0.020	0.012
	TE	0.017	-0.002	0.001	0.005

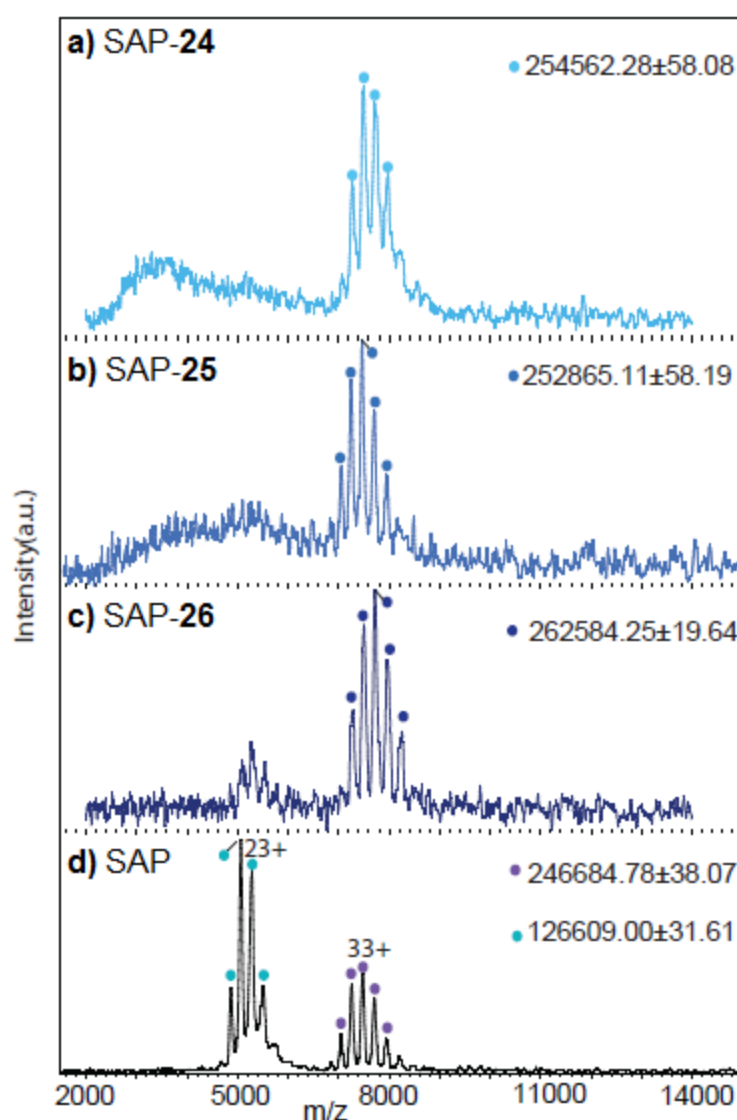
**Table 5.2.** UV-Vis SAP precipitation assay with multivalent ligands 5 minutes after sample preparation. UV-Vis spectroscopy measurements were made of samples of SAP mixed with the multivalent ligands in TC and TE buffer absorption at 320nm wavelength and 5 minutes after sample preparation. Multivalent ligand free samples of SAP in TE and TC buffer were used as control measurements.

Ligand	Buffer	Absorption at 320nm			Average
		Run 1	Run 2	Run 3	
Water Blank	None	0.000	0.000	0.000	0.000
Ligand-free SAP	TC	N/A	N/A	N/A	N/A
S6-5H3	TC	-0.001	-0.002	0.000	-0.003
	TE	-0.001	0.000	0.000	-0.001
S6-5H5	TC	0.001	0.000	0.000	0.000
	TE	-0.004	-0.002	-0.006	-0.004
S6-5H110	TC	0.002	0.007	0.002	0.004
	TE	-0.003	0.000	0.000	-0.001
S6-5H310	TC	0.001	-0.002	0.000	0.000
	TE	0.003	0.001	-0.003	0.000
S6-5H510	TC	-0.001	0.016	0.020	0.012
	TE	0.018	-0.003	0.002	0.006

**Table 5.3.** UV-Vis SAP precipitation assay with multivalent ligands 35 minutes after sample preparation. UV-Vis spectroscopy measurements were made of samples of SAP mixed with the multivalent ligands in TC and TE buffer absorption at 320nm wavelength. The ligand-free SAP samples were not retested as agglutination of precipitated SAP gave poor reproducibility of results.

### 5.3.3. Mass Spectrometry of SAP with decavalent ligands

Mass spectrometry was performed for each of the decavalent ligands to investigate whether the SAP decamer forming capability of the bivalent CPHPC headgroup was retained in the decavalent structures. Samples of SAP were prepared at 0.9 $\mu$ M and incubated with the multivalent ligands at physiological pH prior to the addition of 5.0mM calcium ions. The results pictured in Figure 5.4 show the pentamer/decamer distribution in the absence/presence of the ligand. It is clear from the results that a far greater proportion of decameric SAP was detected in the samples containing the decavalent ligands when compared to the ligand-free sample.



**Figure 5.4.** Whole Protein Mass Spectrometry results of SAP in complex with multivalent ligands a) **24**, b) **25**, c) **26** and d) SAP multivalent ligand free.

Close inspection of the spectra is hampered by the noisy baseline, but it is possible to see some pentameric SAP present in sample ligand **26** (Figure 5.4c). A similar ratio of pentameric/decameric SAP may be present in spectra of ligand **24** (Figure 5.4a) and ligand **25** (Figure 5.4b), but it is not discernible because of the poor signal-to-noise ratio.

The total mass of the protein complexes detected by each experiment varies considerably from the expected mass. Table 5.4 shows the expected mass of each complex and the observed mass. The deviations of the observed values from the

predicted masses are detailed in Table 5.7. The implications and possible reasons for this are discussed in section 5.4.3.

<b>Complex</b>	<b>Ligand Mass</b>	<b>Predicted Pentameric Complex Mass</b>	<b>Predicted Decameric Complex Mass</b>	<b>Observed Mass of Decameric Species</b>
SAP- <b>24</b>	6184	133497 ( $\pm 5.5$ )	260809 ( $\pm 11$ )	254526 ( $\pm 58$ )
SAP- <b>25</b>	6699	134012 ( $\pm 5.5$ )	261324 ( $\pm 11$ )	252856 ( $\pm 58$ )
SAP- <b>26</b>	7149	134462 ( $\pm 5.5$ )	261774 ( $\pm 11$ )	262584 ( $\pm 19$ )
SAP	N/A	127313 ( $\pm 5.5$ )	254625 ( $\pm 11$ )	246684 ( $\pm 32$ )

**Table 5.4.** Comparison of predicted masses for SAP-decameric ligand complexes versus masses of complexes observed in this study. Predicted mass of SAP obtained from Tennent and Pepys (1994).

#### 5.3.4. ITC of SAP with multivalent ligands

Multivalent ligand **21** was synthesised before ligands **22-26** to validate the ligand design's ability to bind to SAP. Therefore, ITC analysis was performed separately from **22-26** and on a different calorimeter to the other multivalent ligands. The ITC of SAP titrated with ligand **21** was performed before the amino acid analysis and therefore before the exact peptide content of the sample was known. Instead it was assumed to be 70% of the dry peptide mass, but subsequent peptide content analysis suggests it may have been significantly lower. It is not possible to recalculate the binding affinity retrospectively with the accurate figure for peptide content because the samples were freeze-dried into sealed vials for amino acid analysis, therefore they may not have contained the same quantity of water as the unsealed vials used to transport **21** for the initial ITC study. The total yield of **21** was exhausted by the time the second ITC runs were performed on the 5 larger ligands and resynthesis was not considered necessary after other multivalent ligands showed significantly stronger binding affinities with SAP that ligand **21** was unlikely to surpass. Therefore, the 1.0 $\mu$ M binding affinity reported for **21** with SAP was interpreted as evidence of binding rather than a precise measurement.

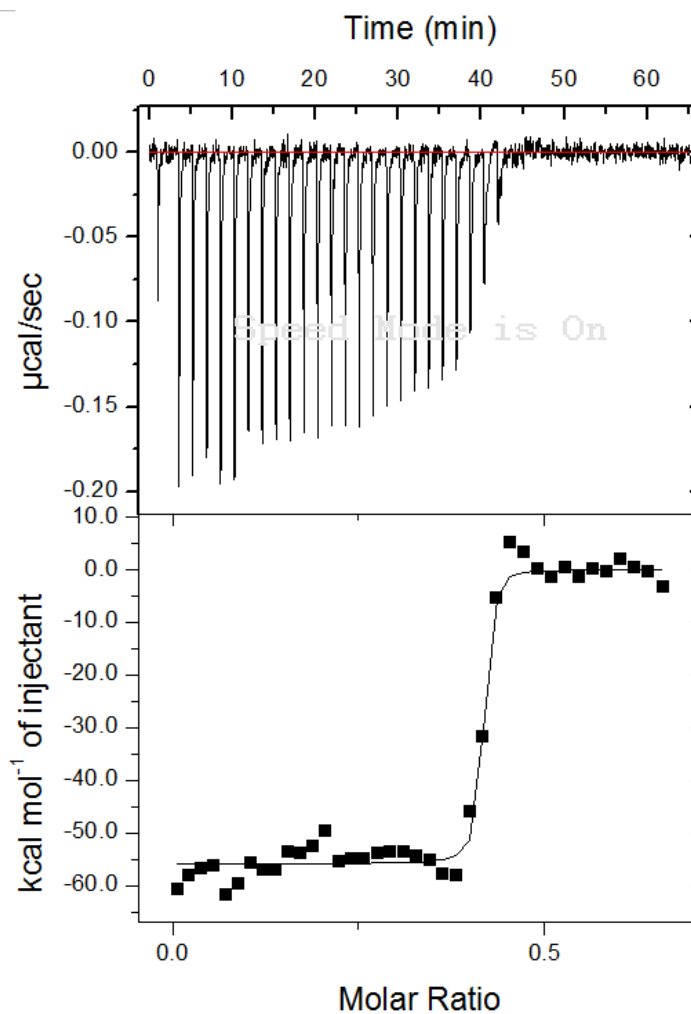
The dissociation constant of **21** was calculated to be almost 20 times stronger than univalent N-acetyl-D-proline (Kolstoe thesis, 2005). This result provided evidence that

the chosen ligand design offered significant promise. **21** was intended to be the smallest pentavalent ligand and it was expected that synthesising larger multivalent ligands would improve binding by ensuring that all SAP binding sites could be occupied by a single ligand simultaneously.

The results obtained from the ITC experiments are listed in Table 5.5, with the results averaged from three identical runs for each ligand (values for all runs can be found in appendix section A.5). All binding isotherms showed an exothermic binding process with negative entropy values indicating a reduction of disorder, both of which were expected. The entropy values were significantly more negative for the decavalent ligands, indicating a greater reduction in disorder and potentially pointing to decamer formation, although this cannot be confirmed by ITC alone.

The strongest binding affinities were recorded for the decavalent ligands, which had dissociation constants approximately two orders of magnitude smaller than the corresponding pentavalent ligand. The strongest binding affinity was observed for ligand **25** with a  $K_d$  of around 250pM. It is crucial to note that this binding affinity is considered too strong for accurate determination by ITC, which is demonstrated by the high standard deviation of  $\pm 200$ pM (Table 5.5). Figure 5.5 shows a typical isotherm for ligand **25** binding to SAP; few points are observed on the transition of the curve making it difficult to accurately determine the dissociation constant. Using the observed  $K_a$  and  $n$  values with a protein concentration of 0.0096mM gives a  $C$  value of over 15000, which is much greater than the desired range of 1-1000 (O'brien *et al.*, 2001). The strongest binding affinities for both the pentavalent and decavalent ligands were recorded for those containing the PEG<sub>3</sub> linker, suggesting that this size is the best balance of size and flexibility for the chosen ligand design.





**Figure 5.5.** Typical results of ITC binding experiment between SAP and multivalent ligand **25**. Top: raw heat change in  $\mu\text{cal s}^{-1}$  per injection. Bottom: Plot of integrated heat change against molar ratio of  $[L]_{\text{total}}/[M]_{\text{total}}$  with sigmoidal line of best fit calculated by non-linear least-squares regression. See appendix section A.6 for binding curves of SAP titrated with multivalent ligands **21-24** and **26**.

<b>Ligand</b>	<b>K<sub>a</sub> (M<sup>-1</sup>)</b>	<b>K<sub>d</sub> (M)</b>	<b>n</b>	<b>ΔH (cal mol<sup>-1</sup>)</b>	<b>ΔS (cal mol<sup>-1</sup>)</b>
<b>22</b>	3.46E7 (±6.48E6)	2.89E-8 (±5.60E-9)	0.741	-2.32E4 (±241)	-43.7
<b>23</b>	2.70E7 (±6.52E6)	3.65E-8 (±9.29E-9)	0.679	-2.61E4 (±403)	-53.3
<b>24</b>	6.96E8 (±2.29E8)	1.44E-9 (±4.74E-10)	0.390	-6.01E4 (±572)	-161
<b>25</b>	3.93E9 (±3.12E9)	2.54E-10 (±2.02E-10)	0.412	-5.17E4 (±572)	-130
<b>26</b>	1.46E9 (7.54E8)	6.85E-10 (±3.54E-10)	0.312	-4.50E4 (±488)	-109

**Table 5.5.** ITC results of multivalent ligands **22-26** binding to SAP.

### 5.3.5. X-ray crystallography of SAP in complex with multivalent ligands

Crystals of SAP were obtained for screens using microlitre drops of 10mg ml<sup>-1</sup> protein and a two-fold molar excess of ligands **23** and **24** mixed with equal volumes of a crystallisation cocktail containing 100mM cadmium chloride, 14-28% PEG 4000 and 100mM sodium acetate, pH 4.6. In both cases, the drops appeared cloudy within 10 minutes of sealing the wells suggesting some amorphous precipitation. The drops remained cloudy throughout, but crystals began appearing within 1 week growing to a maximum size in excess of 1mm x 0.4mm x 0.4mm in 4 weeks.

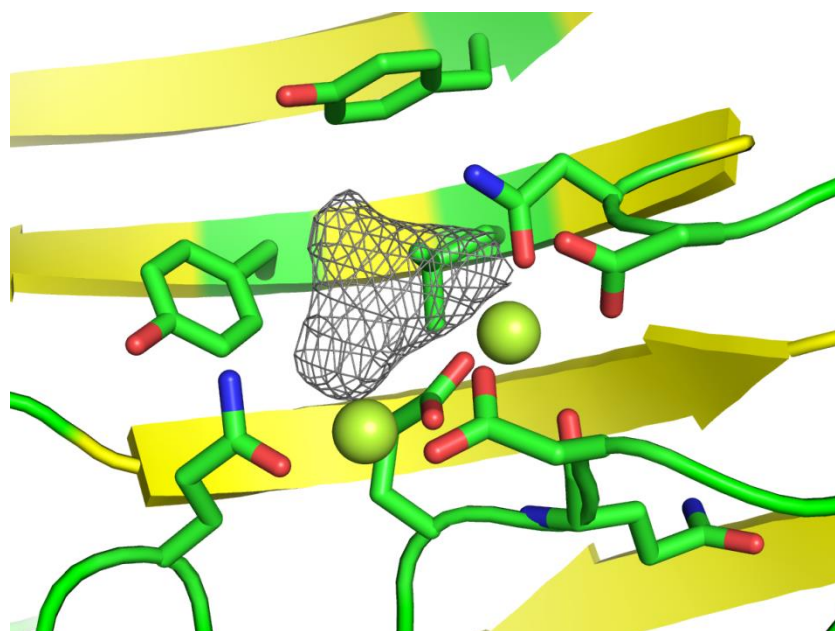
#### 5.3.5.1. X-ray crystallography of SAP-23

The SAP-**23** dataset was processed in spacegroup P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> with unit cell dimensions a = 58.54Å, b = 115.64Å, c = 190.10Å, α = β = γ = 90°. The results of merging and scaling gave a dataset with resolution limit of 2.1Å with outer shell statistics, I/σ(I) of 4.5, completeness 96.6% and R-meas 54.3%.

Molecular replacement was performed on the scaled data using Phaser and PDB file 1SAC as the model to search for a single pentamer in the asymmetric unit as indicated by Matthews (Matthews, 1968) from the CCP4 program suite. Phaser achieved a molecular replacement solution for the dataset with a log-likelihood of 11095 and translation function Z-score of 9.0, indicating a certain solution. The positioning of the model in the PDB file output by Phaser was refined using a single round of rigid body

refinement in Refmac, which gave an initial R-work of 33.29% with an R-free of 32.51%.

The electron density map output by Refmac was viewed in Coot to look for evidence of ligand binding to SAP. The density of the calcium binding site showed two strong spheres of density in calcium binding loop, it was inferred from the crystallisation conditions that these were cadmium ions. Therefore, 10 cadmium ions were built into the model and the structure refined in Refmac. A small Y-shaped portion of electron difference density was observed adjacent to the cadmium binding site in the area where ligands for SAP are observed (Figure 5.6). The electron density was identified as either water or acetate ions, from the ammonium acetate buffer used in the crystallisation cocktail. This indicated that ligand **23** was not occupying the calcium-binding sites of SAP and could not be detected elsewhere within the asymmetric unit.



**Figure 5.6.** Model of the metal-binding site of SAP-**23** constructed from the SAP-**23** crystal data. The Y-shaped difference electron density map (grey mesh,  $1.5\sigma$ ) adjacent to the binding site was identified as being too small to be the expected multivalent ligand, instead it was most likely to be an acetate ion. Two cadmium ions (limon yellow spheres) were identified from the crystallisation conditions bound in the calcium-binding site of SAP. Model colours: cartoon representation:  $\beta$ -sheet (yellow ribbon), loop regions (green ribbon); ball and stick representation: carbon (green), oxygen (red), nitrogen (blue).

The model was subjected to a round of restrained refinement in Refmac to improve the calculated phases and rule out the possibility that the shape of the binding site electron density in the initial map was due to phase error. Subsequent observation of the newly calculated map in Coot showed no improvement in the electron density at the ligand binding site. Therefore, no further effort was made to refine the structure.

### 5.3.5.2. X-ray crystallography of SAP-24

The data was processed using Mosflm in monoclinic spacegroup C2 with unit cell dimensions  $a = 169.28\text{\AA}$ ,  $b = 95.45\text{\AA}$ ,  $c = 88.18\text{\AA}$ ,  $\alpha = 90.00^\circ$ ,  $\beta = 120.29^\circ$  and  $\gamma = 90.00^\circ$ . Monoclinic C-centred unit cell can also be processed in body-centred space group, I2. The generally accepted convention for selecting between C2 and I2 is to select the system that gives the shortest vectors on the  $ac$  plane with  $b$  as the unique axis and the  $\beta$  angle non-acute (Mighell, 2002). The CCP4 program Pointless indicated I2 was the preferred space group for the data with unit cell dimensions  $a = 88.18\text{\AA}$ ,  $b = 95.45\text{\AA}$ ,  $c = 146.20\text{\AA}$ ,  $\alpha = \gamma = 90.00^\circ$  and  $\beta = 91.10^\circ$  and Reindex was used to apply the transformation from C2 to I2. The data was scaled and merged using Scala to a resolution of  $2.02\text{\AA}$ , which gave an  $I/\sigma(I)$  of 2.2, R-meas of 88.0% and 100.0% completeness in the outer shell.

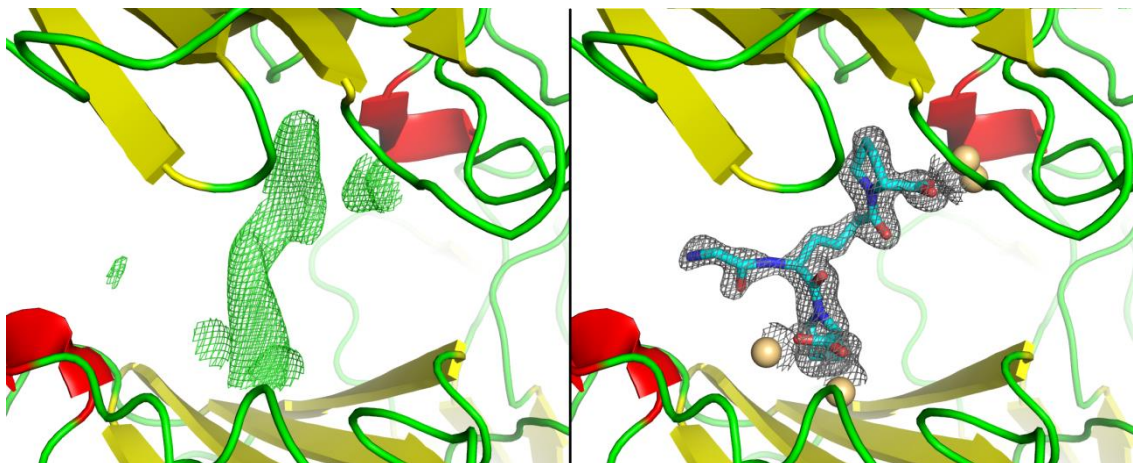
Solution	Spacegroup	Penalty	Unit Cell Dimensions					
			a (Å)	b (Å)	c (Å)	$\alpha$ (°)	$\beta$ (°)	$\gamma$ (°)
1	P1	0	88.1	94.9	97.2	119.2	116.0	90
2	P1	1	88.1	94.9	97.3	60.7	64.0	90
3	C2	3	169.5	95.0	88.2	90.0	120.2	90
4	I222	25	87.9	94.8	146.8	90.0	90.0	90
4AVV*	C2	N/A	154.4	108.6	120.3	90.0	138.5	90

**Table 5.6.** Mosflm unit cell determination results for SAP-24 and unit cell dimensions of PDB file 4AVV.

\*Data taken from [www.rcsb.org](http://www.rcsb.org) on 28/05/2014. Chosen spacegroup and unit cell dimensions of SAP-24 dataset highlighted in blue.

The program Matthews predicted one SAP pentamer in the asymmetric unit, which were searched for using molecular replacement program Phaser with PDB file 1SAC as the search model. Phaser obtained a solution with a LLG of 8440 and TFZ of 8.2, indicating a certain MR result. The molecular replacement solution output by Phaser was initially refined using the rigid body refinement function in Phenix, giving initial R-free of 37.20% and R-work 36.90%. Phenix was then used for all subsequent rounds of restrained refinement applying NCS restraints. The molecular graphics package, Coot was used for model building yielding a final model with R-work, 20.43% and R-free of 23.91%.

A single pentamer was found in the asymmetric unit arranged B-face to B-face with continuous difference electron density visible between the pentamer and a neighbouring pentamer related by the two-fold rotation axis (Figure 5.7). A total of 10 cadmium ions were built into the calcium-binding sites of SAP. In addition, a further 25 cadmium ions were built into areas of strong electron density highlighted by Coot's unmodelled blob validation function, some of which were built with reduced occupancy to highlight the ambiguity in their identity. Five CPHPC-N headgroups of ligand **24** were built into the structure straddling the gap between opposing binding sites in the same manner as observed for CPHPC in PDB file 4AVV. However, there was insufficient continuous electron density to build the entire ligand with adequate certainty, so only the terminal four residues, N-Gly-hGlu(D-Pro)-D-Pro-OH (Gly-N-CPHPC) of each arm of the multivalent ligand were built (Figure 5.7).



**Figure 5.7.** Left: Fo-Fc difference electron density map (green mesh) drawn at  $1.5\sigma$  prior to model building. SAP protomers from two symmetry-related pentamers can be seen in cartoon representation at the top and bottom of the image arranged B-face to B-face. The electron density map shows a continuous region of density not accounted for by the model; further spherical density near the surface of the protomers indicates the presence of metal ions in the calcium binding site of SAP. Right: 2Fo-Fc electron density map (grey mesh,  $1.5\sigma$ ) after model building. Two cadmium ions (cream spheres) and multivalent headgroup Gly-N-CPHPC can be seen tracing the electron density, with well-defined electron density in the chain linking the two D-proline termini indicating a preferred conformation. No further electron density is visible at the N-terminus of the Gly preventing model building and suggesting greater ligand flexibility. Model colours: SAP cartoon representation:  $\beta$ -sheet (yellow ribbon),  $\alpha$ -helical-like regions (red ribbon), loop regions (green ribbon). Ligand atom colours: carbon (turquoise), nitrogen (blue), oxygen (red).

In addition to the ligands and metal ions built into the model, 5 N-acetyl glucosamine (NAG) units were placed in covalent bonding distance to the amide side chain of Asn32, identified as a glycosylation site by Tennent and Pepys (1994). The resolution of the structure also allowed 390 waters and 4 acetate molecules to be built into the electron density.

Space Group	I2
Unit Cell Dimensions	a = 88.18Å, b = 95.45Å, c = 146.20Å, $\alpha =$
Resolution Range	80.04 – 2.02Å
No. of Reflections	481 261
No. of Unique Reflections	79 403
Multiplicity	6.1 (6.1)
Completeness (%)	100.0 (100.0)
R <sub>meas</sub>	0.123 (0.880)
(I)/ $\sigma$ (I)	9.1 (2.2)
<b>Refinement</b>	
Molecular Replacement Model	1SAC
No. of Residues	1020
No. of Solvent Molecules	390 (Water), 25 (Cd), 4 (ACY)
No. of Ligand Molecules	5 (Gly-N-CPHPC), 10 (Cd), 5 (NAG)
Rmsd Bonds	0.01Å
Rmsd Angles	1.25°
Residues in Favoured Regions (%)	97.80%
Residues in Disallowed Regions (%)	0.00%
<b>Average B-factor</b>	40.24
Protein	39.75
Ligand	49.42
Solvent	44.86
R-factor (%)	20.43
R-free (%)	23.91

**Table 5.7.** Processing and refinement statistics for SAP-24 crystal structure. Figures in parenthesis denote results for the highest resolution shell (2.13-2.02 Å).

## **5.4. Discussion**

### **5.4.1. Amino acid analysis**

Accurate peptide content analysis proved to be essential as the values were significantly below 100% and varied widely from the 70% assumption that was made for ligand **21** prior to ITC analysis (Table 5.1). This data shows the importance of accurate quantitation of peptide content prior to use in assays that require accurate knowledge of ligand concentration, such as ITC. The figures are considered to be in an acceptable range, although the 46% for ligand **21** is lower than anticipated. The possible explanation for the lower content is that it was the first ligand to be synthesised and as such the synthetic strategy was still being developed, this could have led to an increase in low molecular weight organic impurities in the final product.

### **5.4.2. Ultraviolet/Visible absorption SAP precipitation assay**

The UV/Vis SAP precipitation assay quickly identified binding of all six multivalent ligands to SAP. The lack of precipitation observed for all samples indicated that they all bound with sufficient affinity to prevent auto-aggregation; there is little more that can be concluded from this assay, but it provided confirmation that further study was warranted. Such an assay should be employed as the first step in all SAP ligand binding investigations.

### **5.4.3. Mass spectrometry of SAP-multivalent ligand complexes**

From the outset of the investigation, there was considered to be a low probability of producing crystals of SAP in complex with the multivalent ligands within the time-frame of the investigation. The large size of the ligands was expected to complicate SAP crystal packing, meaning conditions that had previously produced x-ray diffraction-quality crystals of SAP in complex with univalent ligands were unlikely to be successful with the much larger multivalent ligands. Therefore, mass spectrometry was identified as a suitable alternative strategy to investigate the effect of the decavalent ligands on the distribution of pentameric/decameric SAP in the presence of calcium. In reality, crystallisation of SAP with decameric ligand **24** occurred rapidly and produced suitable crystals for x-ray diffraction. So the mass spectrometry results serve to validate the x-



ray diffraction results for SAP-**24** and provide insight into the nature of the interaction of SAP with ligands **25** and **26**, which have not been crystallised as of yet.

The addition of decavalent ligands to a solution of SAP at physiological pH results in a shift from mostly pentameric SAP to the decameric form being the only species present in detectable quantities (Figure 5.4). However, the observed masses of the complexes, as calculated by maximum entropy, vary from the expected value, calculated by summing the molecular weight of SAP measured by Tennent and Pepys (1994) and the molecular weight of the ligands measured by mass spectrometry in Chapter 4. Table 5.8 shows the deviation of the measured masses from the expected values given by two measurements of the mass of SAP: firstly the mass of SAP calculated by Tennent and Pepys (1994) and secondly from the mass of the pentraxin observed in the ligand-free SAP decamer (Figure 5.4d).

<b>Complex</b>	<b>Complex Mass Predicted from Literature Data</b>	<b>Complex Mass Predicted from Observed Data</b>	<b>Mass Observed by MS</b>	<b>Deviation from Literature Prediction</b>	<b>Deviation from Observed Prediction</b>
SAP- <b>24</b>	260809 ( $\pm 11$ )	252868 ( $\pm 32$ )	254526 ( $\pm 58$ )	-6283	+1658
SAP- <b>25</b>	261324 ( $\pm 11$ )	253383 ( $\pm 32$ )	252856 ( $\pm 58$ )	-8468	-527
SAP- <b>26</b>	261774 ( $\pm 11$ )	253833 ( $\pm 32$ )	262584 ( $\pm 19$ )	+810	+8751
Ligand-free SAP	254625 ( $\pm 11$ )	246684 ( $\pm 32$ )	246684 ( $\pm 32$ )	-7941	N/A

**Table 5.8.** Deviation of masses of SAP-decavalent ligand complexes from the expected values. The literature value for mass of an SAP decamer containing two pentamers and a single multivalent ligand was calculated using the mass of SAP from Tennent and Pepys, 1994. The predicted value calculated by adding the mass of the ligand measured by MS in Chapter 4 and the observed ligand-free mass of decameric SAP (Figure 5.4d). All figures are in units of Daltons (Da).

Table 5.8 shows two different calculated values for the SAP-ligand complexes. The first are calculated using the ligand molecular weights measured in Chapter 4 and the reported mass of SAP from Tennent and Pepys (1994). The second are the values calculated using the ligand molecular weights from Chapter 4 and the observed mass of the ligand free decamer from Figure 5.4d. Each of the observed masses of the SAP-ligand decamers (Figure 5.4a-c) differs significantly from one of the calculated values, in some cases the difference is almost 8000Da. Furthermore, the value for the ligand-free decameric complex is less than would be expected based on the mass of the pentameric SAP detected in the same experiment (Figure 5.4d). The observed mass for pentameric SAP deviates by only around 700Da from the expected literature value. This suggests that deviation in the decamer mass may be the result of an error in the maximum entropy calculation of the decameric complex. However, the exact cause of the deviations is difficult to pinpoint from this study. Van Dongen and Heck investigated the use of ESI-MS to study non-covalent proteins interactions, using the example of Apo-Concanavalin A binding with selected carbohydrates ligands (2000). They concluded that a number of factors affected the quality of results and that each new non-covalent interaction study necessitates a period of equipment and sample optimisation to produce highly precise data.

The quantity of multivalent ligands produced and the availability of protein for this investigation simply did not allow for the period of optimisation suggested by Van Dongen and Heck (2000). This does not invalidate the results, but it does add some uncertainty as to the nature of the protein-ligand interaction and the identity of the complexes detected by MS. The variability in the recorded mass means it is not possible to infer whether the decameric complexes contain the multivalent ligands or whether they induce complex formation without involvement in the final arrangement. However, with unequivocal x-ray diffraction data showing non-covalent cross-linking of two SAP pentamers by ligand **24**, it is highly likely that the increased proportion of SAP decamers observed by mass spectrometry is due to the presence of the same SAP-multivalent ligand complexes observed by x-ray crystallography.

Like so much early stage research the mass spectrometry data presented in this study offers tantalising results that require optimisation to increase precision. The decavalent ligands have been shown to induce decamer formation in SAP by mass spectrometry, which is an early indication that the non-covalent cross-linking ability of CPHPC has been preserved in the decavalent N-CPHPC headgroups synthesised in this investigation. Although the reliability of these results is questionable when viewed in isolation, when taken in combination with the reaction stoichiometry calculated by ITC and the decameric SAP-**24** complex observed by x-ray diffraction, these results add an acceptable level of confidence that all the decavalent ligands are non-covalently cross-linking SAP pentamers in the same manner as bivalent ligand CPHPC, but with potentially stronger binding affinities.

#### **5.4.3.1. Optimising mass spectrometry for SAP-decavalent ligand complexes**

Any future mass spectrometry of SAP-multivalent ligand complexes should try to achieve sharper peaks for the desired product and a greater signal-to-noise ratio. At least part of the mass variability and poor baseline may be the result of the ligands being present as chloride salts. Chloride ions suppress ionisation and can lead to excessive adduct formation during mass spectrometry experiments (Henández and Robinson, 2007). Adduct formation will result in peak widening and this makes deconvolution of the mass more difficult. In future analysis it would be best to perform the buffer exchange with a higher concentration of ammonium acetate to improve the precision of the mass determination. A buffer concentration of 1M may be appropriate, as this has been shown to improve peak resolution and reduce noise at the low mass range (Verkerk and Kobarle, 2005). Acetate ions do have the potential to bind to SAP, so higher concentrations could prevent complex formation between SAP and multivalent species. This would need to be carefully monitored, but given the strong binding avidity of the multivalent ligands to SAP it would be hoped that acetate would not disrupt complex formation.

It is also conceivable that the mass may have varied due to fragmentation. These fragments may be hidden by or even contributing to the noisy baseline in the loss mass range. If fragmentation is occurring due to harsh ionisation conditions then a lower cone

voltage should increase the observed mass of the complex.

#### **5.4.3.2. Mass spectrometry of pentavalent ligands 21-23 in complex with SAP**

No effort was made to perform mass spectrometry with SAP and the pentavalent ligands **21-23**. The potential for **21-23** to form decameric complexes with SAP was considered low. This was based on the fact that once a single headgroup has bound to a pentamer, the probability of the other headgroups coordinating the other calcium-binding sites on the same pentamer is more likely than binding another incoming pentamer. ITC observations offered further evidence that decamers do not form between SAP and **21-23**. The stoichiometry of the binding is roughly double that observed for the decameric ligands and much closer to ITC observations of univalent ligand binding to SAP made by Kolstoe (Kolstoe thesis, 2005). Furthermore, the entropy of the observed ITC binding for pentavalent ligands is significantly less negative than that of the decavalent ligands, indicating a more disordered system, which suggests **21-23** form pentameric complexes with SAP rather than decamers. For these reasons, mass spectrometry of the pentameric ligands with SAP was not considered an appropriate use of the limited protein stock available.

#### **5.4.4. Isothermal titration calorimetry**

The initial ITC investigation between SAP and **21** showed an estimated binding affinity 20-fold greater than univalent N-acetyl-D-proline (NADPro; Pepys *et al.*, 2002). **21** may not have been capable of binding to all SAP binding sites on a single pentamer simultaneously because it had the shortest PEG chain in the linkers used to bridge the central scaffold to the headgroups and therefore should have the smallest radius of all the pentavalent ligands. If **21** can only partially occupy the binding sites of an SAP pentamer then there would be a significant penalty to the binding affinity from the high flexibility of the ligand. The increased affinity over univalent NADPro provided confidence that the methodology and ligand structure had the potential to produce ligands with significantly tighter binding avidity if they had a larger effective radius to enabling coordination of all five SAP binding sites simultaneously.

The fact that the exact peptide content of the ligand had not been determined prior to the ITC reduces the precision of the result. However, larger multivalent ligands were shown to have significantly stronger binding affinities with SAP, so an exact measurement was deemed unlikely to significantly alter the results of the study.

#### **5.4.4.1. Isothermal titration calorimetry of ligands 22-26**

The second ITC investigation utilised a different calorimeter that was vastly more efficient with the protein stock available, but still provided heat changes well in excess of the detection limit of  $0.1 \mu\text{cal s}^{-1}$  (MicroCal ITC 200, User Manual, GE Life Sciences, Uppsala, Sweden). The ITC results show the decavalent ligands to be around 100-fold tighter binders to SAP than their pentavalent counterparts (Table 5.5). The values calculated for the dissociation constant are beyond the accurate measurable limit of the apparatus, which is a promising result, but will require a different binding assay to measure accurately (see section 5.4.4.3).

#### **5.4.4.2. Stoichiometry deviation from expected values**

The stoichiometry,  $n$ , of the binding is below the expected values of 1 for the pentavalent ligands and 0.5 for the decavalent ligands (Table 5.5). In ITC, calculated stoichiometry is known to vary when the sample purity is less than 99% (Giancola and Pagano, 2012). The purity of the peptide content of the multivalent ligands was found to be between 85-96% in Chapter 4 as determined by LCMS (Table 4.2). This will have increased the error in the calculation of reaction stoichiometry. As this is the first reported case of synthesis and design of multivalent ligands for SAP, it is inevitable that the quality of the products is lower than compounds produced by established methodologies, such as commercial products obtained from chemical suppliers. The protein concentration may also have deviated from the expected value, SAP auto-aggregation in the sample cell prior to addition of the ligand may have been the cause of the error. Aggregated SAP would not be available for binding and therefore saturation would be reached at an artificially lower concentration of protein, resulting in a lower stoichiometry value. It was observed that the reference power began to vary slightly from the specified  $6 \mu\text{cal s}^{-1}$  value between runs, which can be symptomatic of a dirty sample cell (MicroCal ITC 200, User Manual, GE Life Sciences, Uppsala, Sweden).

This may indicate that some SAP had precipitated and coated the sample cell. However, no precipitation was observed in the solution removed from the ITC at the end of any titration.

It is worth noting that Kolstoe found an  $n$  value of 0.78 for univalent ligand phosphoethanolamine, PE, and 0.36 for bivalent ligand, CPHPC (Kolstoe thesis, 2005). X-ray diffraction has shown PE to interact with a single subunit of SAP in a 1:1 binding stoichiometry (Mikolajek *et al.*, 2011) and CPHPC is known to interact with two subunits from different SAP pentamers in a 1:2 ratio (Pepys *et al.*, 2002). Therefore, the  $n$  values for the multivalent ligands are not too dissimilar to values observed in other SAP ITC experiments. It is realistic to conclude from the calculated stoichiometry that the pentavalent ligands bind a single SAP pentamer in a 1:1 ratio and the decavalent ligands interact with two SAP pentamers in a 1:2 ratio. This inference is in agreement with the decameric complex of SAP-**24** observed by x-ray diffraction.

#### **5.4.4.3. Binding cooperativity**

The reason for synthesising multiple D-proline headgroups into a single molecule was to take advantage of the vastly increased avidity from cooperative binding (Boas and Heegaard, 2004). Cooperative binding is the increase in binding activity achieved in a multivalent species over the corresponding univalent components working independently. The binding affinity enhancement can be attributed one of two effects: the additive effect or the cluster effect. Additive binding is the binding efficiency increase achieved by having a higher number of binding entities per molecule in a multivalent ligand than in a univalent ligand. The cluster effect, also known as the dendritic effect, is observed when synergies of simultaneous binding of multiple ligands results in a larger increase in binding activity than would be expected by a solely additive process. Cooperative binding can be positive and enhance the affinity of each binding event resulting in an equilibrium constant greater than the additive equilibrium constant. Alternatively, the binding can be negatively cooperative, the affinity of binding is less than expected for a purely an additive cooperative process.

In multivalent ITC experiments, a sigmoidal curve is indicative of a cooperative binding

process (Schön and Freire, 1989). Sigmoidal curves were observed for all ITC experiments of multivalent ligands **21-26** binding to SAP. The sharp transition from unsaturation to saturation of SAP binding sites makes the sigmoidal shape appear almost like a step transition, showing that the cooperative effect has greatly increased the affinity of NADPro binding to SAP.

In Biochemistry, the extent of cooperativity in a multivalent binding process can be determined using the  $\alpha$ -value, which is calculated by Equation 5.4 (Mammen *et al.*, 1998).

$$\alpha = \frac{\log K_{multi}}{\log(K_{uni})^n} \quad \text{Equation 5.4.}$$

In Equation 5.4,  $K_{multi}$  is the association constant of the multivalent process,  $K_{uni}$ , the association constant for the univalent binding and  $n$  the number of binding events in the multivalent process. Theoretically, if a process is positively cooperative then  $\alpha > 1$ , additive:  $\alpha = 1$  or negatively:  $\alpha < 1$  cooperative. However, very few processes are determined to be positively cooperative by this measure (Ercolani and Sciaffino, 2011). So the  $\alpha$ -value is more instructive for comparing the relative cooperativity of different multivalent binding events. In this case, for comparing cooperativity of SAP binding to bivalent ligand CPHPC and penta- and decavalent ligands **21-26**.

The  $\alpha$ -values for the multivalent ligands were calculated using the dissociation constant of 15  $\mu$ M for univalent N-acetyl-D-proline reported by Pepys *et al.* (2002); the reciprocal of the dissociation constant being the association constant. The  $\alpha$ -values for all the multivalent ligands are less than 1, indicating negative cooperative binding (Table 5.9). The Log  $K_a$  calculated for NADPro was multiplied by an  $n$  value of 5 to calculate the  $\alpha$ -value for pentavalent ligands **21-23** and an  $n$  value of 10 was used for calculating the  $\alpha$ -value for decavalent ligands **24-26**.

<b>Ligand</b>	<b>Log K<sub>a</sub></b>	<b><math>\alpha</math></b>
N-acetyl-D-proline	4.82	1.00
<b>21</b>	6.00	0.25
<b>22</b>	7.54	0.31
<b>23</b>	7.43	0.31
<b>24</b>	8.80	0.18
<b>25</b>	9.59	0.20
<b>26</b>	9.16	0.19
CPHPC	8	0.83

**Table 5.9.** The  $\alpha$ -values for the binding constants of the multivalent ligands binding to SAP as measured by ITC. Ligands **21-23** are compared to the additive binding constant predicted for 5 NADPro, while **24-26** are compared to the additive constant predicted for 10 NADPro binding to SAP.

The complex structure of the ligands means that there is significant freedom of movement and many different conformations in which they can exist. Binding to a macromolecule will significantly reduce the freedom of movement, thus greatly increasing the negativity of the entropy term. The Van't Hoff equation shows how a more negative entropy term weakens the equilibrium constant for a binding process (Equation 5.5).

$$\ln K = -\frac{\Delta H}{RT} + \frac{T\Delta S}{R} \quad \text{Equation 5.5.}$$

The greater structural complexity of multivalent ligands over univalent NADPro slightly reduces the binding affinity for the multivalent process and that is why the values of  $\alpha$  are less than 1. This result is not unexpected, as the vast majority of interactions measured by this ratio appear to be negatively cooperative (Ercolani and Sciaffino, 2011). Performing the same calculation for bivalent ligand CPHPC ( $\log(K_{bi}) / \log(K_{uni})^2$ ), based on the dissociation constant of 10nM gives a value for  $\alpha$  of 0.83, which shows the binding of CPHPC to SAP to be more cooperative than the multivalent ligands in this study, even though it has a weaker binding affinity. The higher  $\alpha$ -value for CPHPC reflects the fact that the palindromic structure of the molecule allows a mode of binding that is impossible for univalent NADPro, namely non-covalent



decamerisation of two pentamers. The decameric complex of SAP-5(CPHPC)-SAP can be visualised as each pentamer bound to a pseudo pentavalent ligand that presents five NADPro headgroups in perfect geometry for binding. Furthermore, the five points of non-covalent attachment decrease the likelihood of the decamer disassociating, so any headgroups that do unbind from the calcium-binding site are likely to quickly rebind as they are kept in close proximity to the binding site by the decameric arrangement. In contrast, a univalent ligand, such as NADPro, is more likely to dissociate from the binding site of SAP and not rebind as there is nothing preventing it diffusing into the surrounding solvent. Pentavalent ligands **21-23** overcome this issue by binding to SAP at five identical sites, if any one headgroup dissociates it is kept in close proximity to the binding site by the remaining bound ligands to which it is covalently tethered. The lower dissociation constant is therefore derived from the fact that the probability of all five ligands dissociating is significantly lower than a single univalent NADPro dissociating.

The dissociation constant measured for the decavalent ligands was slightly smaller than observed for CPHPC. The reason for this is the same reason the pentavalent ligands bind more strongly than univalent NADPro. The probability of ten decavalent headgroups dissociating is less than the probability of the two headgroups in CPHPC dissociating and diffusing out of the decameric complex. The short tether linking the two NADPro in CPHPC and the mechanism which uses another SAP pentamer as a pseudo-scaffold is clearly a highly efficient binding mechanism and although the affinity is slightly improved by tethering five CPHPC together it does not result in the same increase in affinity over CPHPC that CPHPC has over NADPro.

#### **5.4.4.4. Accurate determination of the dissociation constant**

Decavalent ligands **25** and **26** have been estimated to bind to SAP with a sub-nanomolar dissociation constant, which is considered to be beyond the measurable range of the ITC. Therefore, an alternative method of measuring the dissociation constant should be sought. Typically in this situation a competition assay would be employed (Sigurskjold, 2000), where a weak binding ligand is bound to SAP and then replaced upon addition of the stronger binding multivalent ligand. The energy required to break the interaction of

the weak binding ligand reduces the measured affinity of the stronger binding process, ideally bringing the dissociation constant into the measurable micromolar-nanomolar range. However, this methodology is not appropriate for this investigation because the weak binding ligand would need to have a dissociation constant in the high micromolar range, perhaps around 100 $\mu$ M. The only known ligands for SAP that have this affinity are univalent or bivalent. It would be impossible to determine if all five univalent ligands had been displaced by the multivalent ligand in a competition assay. Therefore, it would necessitate the availability of a weak binding multivalent ligand for SAP for the analysis to be a fair and interpretable competition assay and no such ligand is known.

An alternative strategy to calorimetry would be a fluorescence assay. This method requires prior knowledge of the binding stoichiometry to calculate the dissociation constant from the concentration of unbound ligand as ligand is slowly added to a known concentration of protein. From the ITC and crystallography experiments is known with reasonable confidence to be 1:1 for SAP with pentavalent ligands and 2:1 with decavalent ligands. From this method it will be possible to produce a Scatchard plot which shows the dissociation constant from the gradient (Voet, 1995) and the Hill equation to determine the extent of cooperative binding.

Often fluorescence assays require one of the reactants to be labelled with a dye whose fluorescence changes upon binding (Siegel and Linstad, 2010). This may not be necessary because the binding of NADPro headgroups to SAP involves close packing of the pyrrolidine ring against two tyrosine side chains. This packing may quench the natural UV fluorescence of the phenol side chains sufficiently to be used to quantitatively assess the extent of binding. This should be investigated prior to any modification of either SAP or the multivalent ligands.

If however a fluorescent dye is required, then the C-terminal carboxyl of the multivalent ligand could be used as currently this is only used to link the peptide to the solid phase during synthesis. The most common fluorescent dyes utilise amine or thiol functional groups for attachment to a polypeptide. Post-cleavage modification of the C-terminus of

the central scaffold of the multivalent ligands with diaminoethane would be a simple way to place a free amine on the molecule. This would require a slightly different protecting group strategy, but it could easily be achieved without significantly changing the synthetic strategy.

#### **5.4.5. X-ray crystallography**

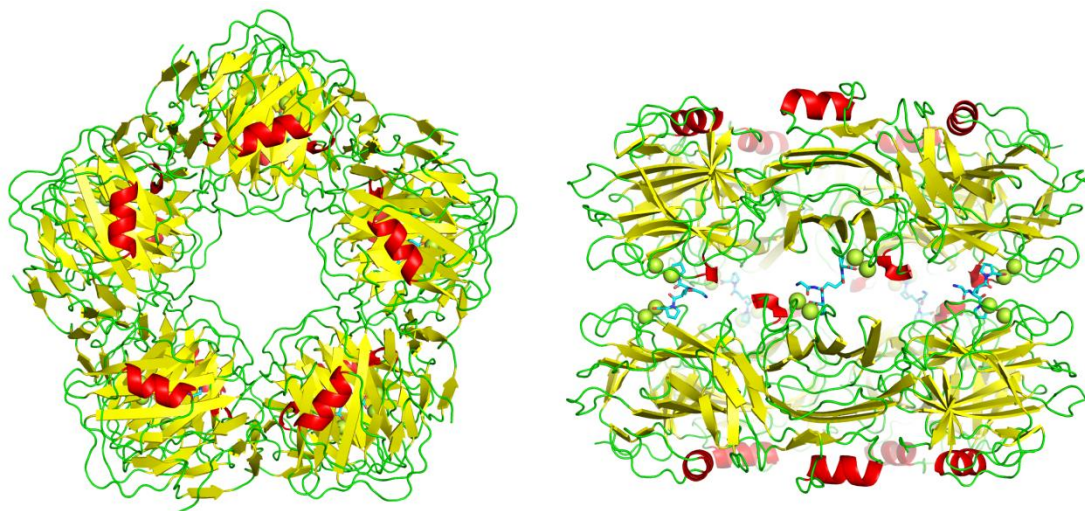
The crystal structure of SAP co-crystallised with **24** provides conclusive evidence of the ligand's capability to induce decamer formation by non-covalent cross-linking of two SAP pentamers in the same manner as observed for CPHPC (Figure 5.8; Kolstoe *et al.*, 2014). In contrast, it has not been possible to crystallise SAP in complex with multivalent ligand **23**, despite obtaining crystals from SAP-**23** crystal screens in the same conditions as those used to crystallise SAP with **24**.

Addition of divalent cations to crystallisation cocktails is a recognised method for bridging non-covalent interactions between protein molecules (McPherson, 1982; McPherson, 1990); cadmium ions are particularly known for improving crystallisation in this manner (Trakhanov *et al.*, 2008). While this has been shown to be beneficial to the diffraction-quality of crystals produced for decameric SAP (Kolstoe *et al.*, 2014), it may have been detrimental in the preparation of SAP-**23** crystals. Multiple contacts between pentamers bridged by cadmium ions created smaller channels between pentamers leaving insufficient space for ligand **23**. In place of the desired ligand, acetate ions from the ammonium acetate buffer can be seen completing the coordination spheres of the metal ions in the calcium-binding site of SAP (Figure 5.6).

##### **5.4.5.2. SAP-24 complex structure analysis**

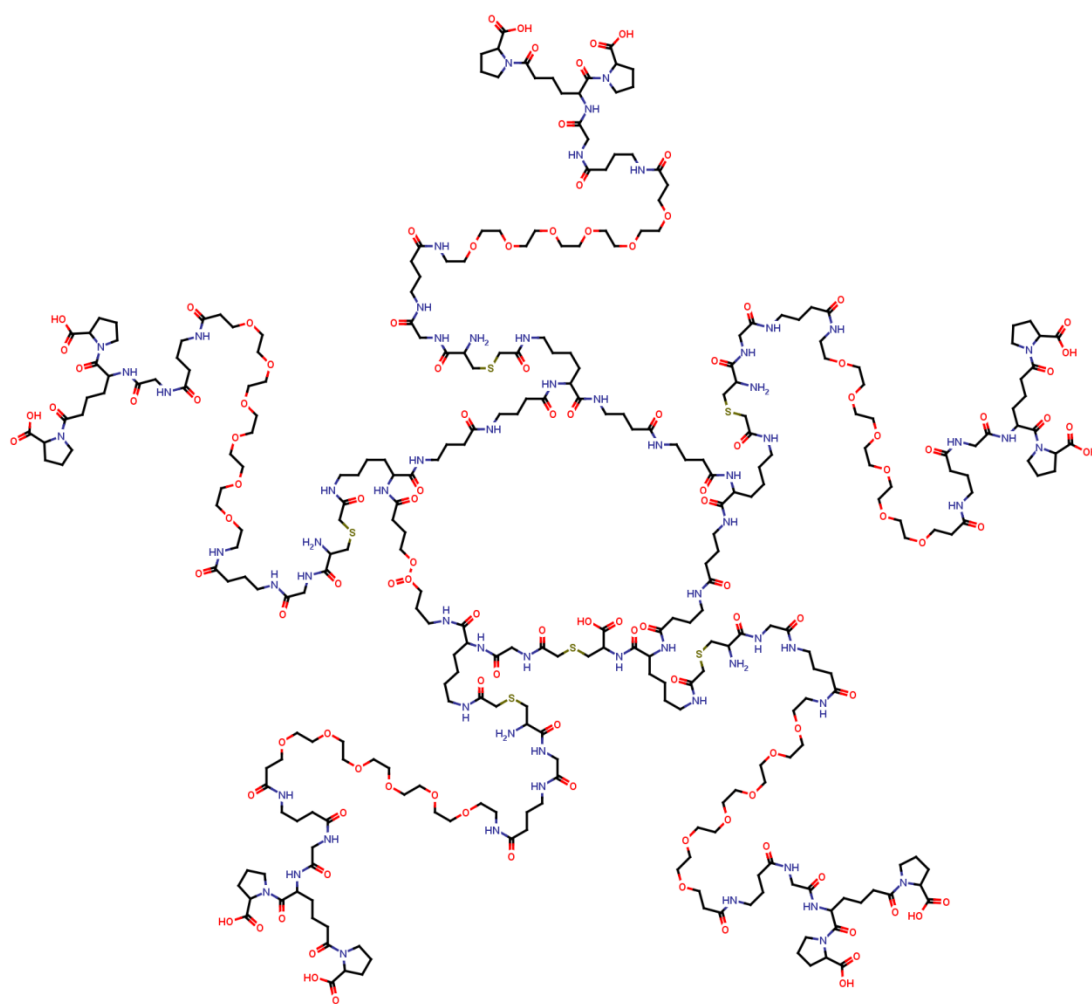
Both pentamers in the SAP-**24** complex displayed the characteristic features of hSAP described by Emsley *et al.* (1994) and detailed in Chapters 1 and 2 (Figure 5.8). A continuous area of difference density was visible running between calcium-binding sites from each pair of opposing protomers (Figure 5.7). This was identified as the N-CPHPC headgroup from multivalent ligand **24**. The continuous nature of the density indicated that N-CPHPC interacts with two protomers from opposing pentamers simultaneously. Therefore, the non-covalent, cross-linking binding mechanism of CPHPC has been

preserved in the modified N-CPHPC headgroup used in the decavalent ligands.



**Figure 5.8.** Cartoon representation for the SAP-**24** crystal structure viewed top down (left) and side on (right) generated with Pymol. 2 SAP pentamers are non-covalently cross-linked by **24** forming a B-face to B-face decamer. Only the 5 N-CPHPC headgroups of **24** could be built into the electron density. Model colours: cartoon:  $\alpha$ -helix (red ribbon),  $\beta$ -sheet (yellow ribbon), loop regions (green ribbon); ball and stick representation: carbon (turquoise), nitrogen (blue), oxygen (red) and cadmium (limon yellow spheres).

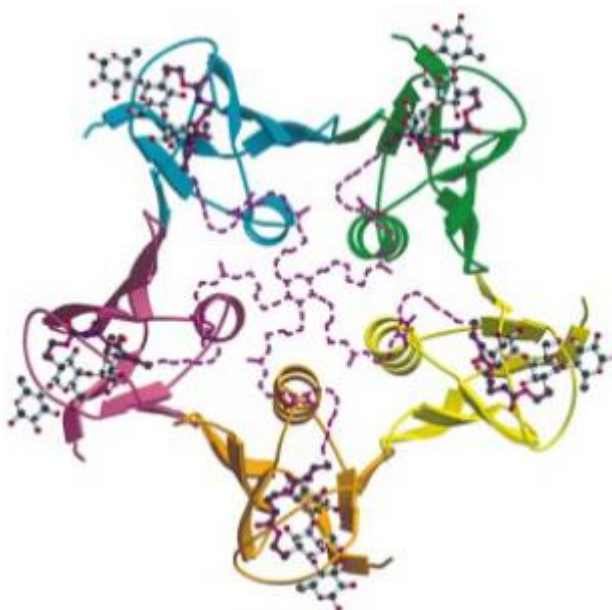
Once the N-CPHPC headgroup had been built into the map, further electron density was visible protruding adjacent to the N-terminus of N-CPHPC. This was determined to be the Gly residue used to ligate the headgroup to the rest of the multivalent ligand. No further electron density for the ligand could be conclusively identified. This implies that the central cyclic core and polyethylene glycol linkers have some freedom of movement and no one particular conformation dominates in the time frame of the experiment. Furthermore, it leaves a large portion of the ligand undetermined, Figure 5.9 shows the complete structure of ligand **24**. Only the terminal four residues of each arm have been located in the electron density map for the SAP-**24** crystal structure.



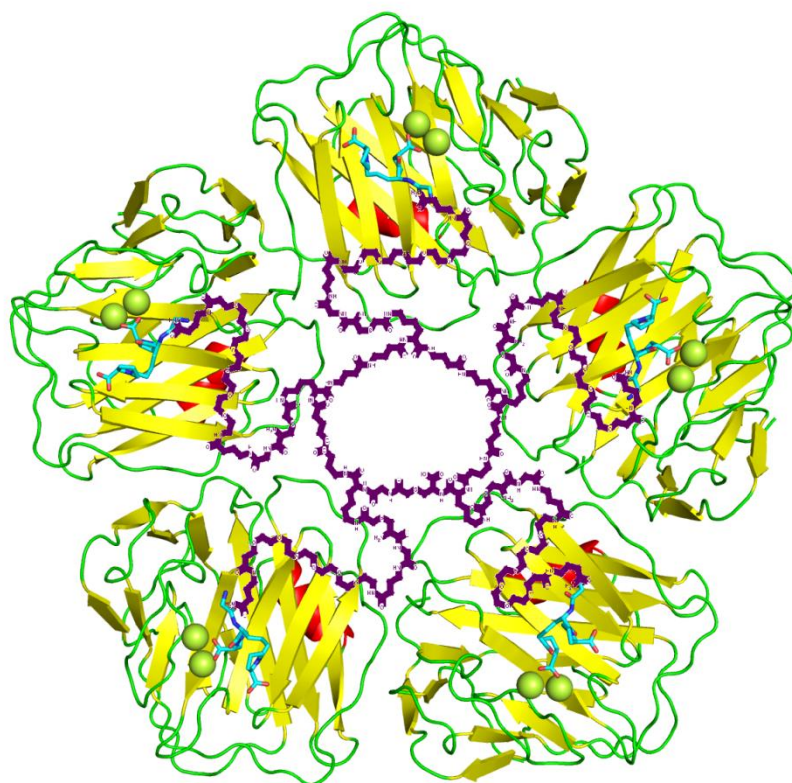
**Figure 5.9.** Structure of ligand **24**. Image drawn with Marvin Sketch (ChemAxon, Budapest, Hungary).

The lack of electron density means it is difficult to determine the binding stoichiometry. When there is not continuous electron density linking all the headgroups together it is difficult to conclude that all the headgroups belong to a single molecule of **24**. Conceivably, it would be possible for multiple ligand molecules to bridge contacts between multiple decamers forming a non-covalent polymer of SAP decamers. However, from the electron density that has been modelled, all Gly residues are arranged with their N-termini pointing towards the central pore of the complex, so the implication is that the headgroups belong to a single multivalent ligand molecule. Similar disorder in the central scaffold was observed by Kitov *et al.* (2000) in the crystal structure of their STARFISH ligand in complex with Shiga-like toxin 1. Kitov *et al.* produced a graphic plotting the position they expected the ligand to occupy in the

complex Figure 5.10) and a similar situation is expected for the SAP-**24** structure is presented here (Figure 5.11).



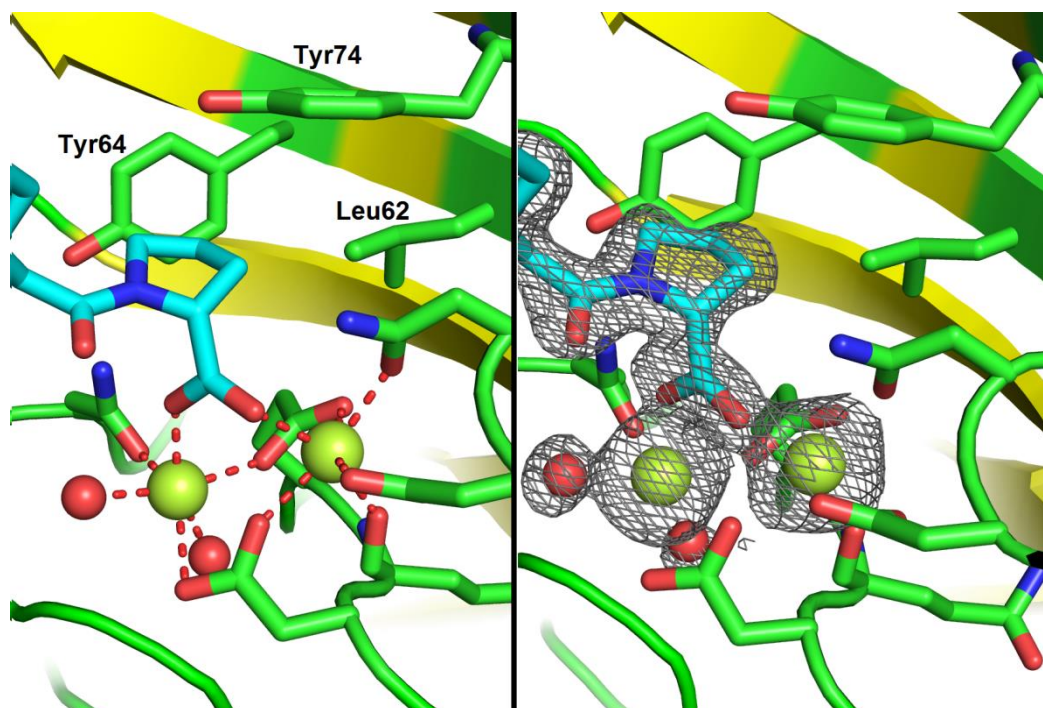
**Figure 5.10.** Ball and stick model of STARFISH Shiga-like toxin inhibitor binding to SLT-1 (cartoon representation) developed by Kitov *et al.*, 2000. The authors were unable to reliably build the entire structure of the inhibitor in their crystal structure due to its flexibility. Therefore, they proposed a possible arrangement for the central core of the ligand (magenta dotted line). It is likely that ligand **24** is similarly arranged in the crystal structure of SAP-**24**.



**Figure 5.11.** View of the A-face of a single SAP pentamer in the crystal structure of SAP-**24** with a proposed position for the central core and polyethylene glycol linkers of ligand **24** (purple chain). As no defined conformation for the ligand could be discerned beyond the Gly-N-CPHPC headgroups, the proposed position should be taken as only a possible conformation that illustrates the relative size of the ligand to the protein. Model colours: cartoon:  $\alpha$ -helix (red ribbon),  $\beta$ -sheet (yellow ribbon), loop regions (green ribbon); ball and stick representation: carbon (turquoise), nitrogen (blue), oxygen (red) and cadmium (limon yellow spheres).

The D-proline termini of the headgroup can be seen coordinating the cadmium ions in the calcium-binding site of SAP through their carboxyl groups in the same manner as seen for D-proline in Chapter 2 and as for CPHPC described by Kolstoe *et al.* (2014). The pyrrolidine rings of D-proline fit into SAP's hydrophobic pocket formed by Leu62, Tyr64 and Tyr74 (Figure 5.12).



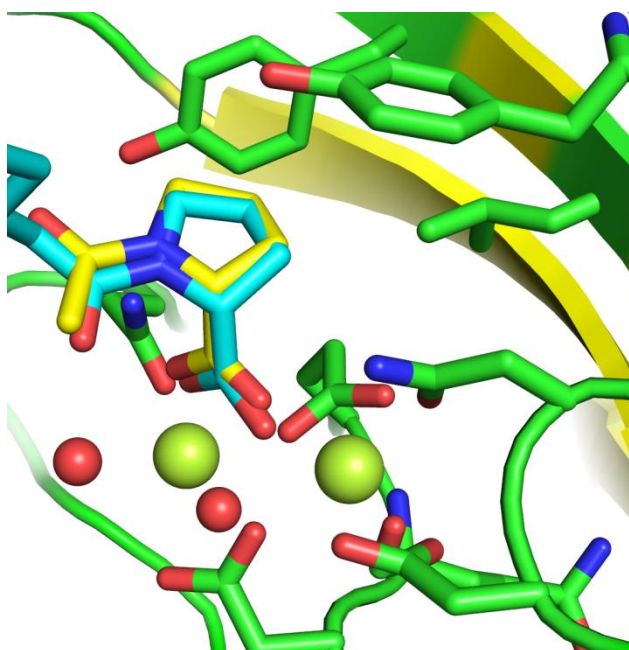


**Figure 5.12.** Calcium-binding site of SAP in SAP-**24** crystal structure. Left: the cadmium ions are coordinated by the side chains of multiple SAP residues and water molecules (red dashed lines). The coordination spheres are completed by the carboxyl of a single D-proline from the N-CPHPC headgroup of **24**. Right: the fit of the coordinated cadmium ions, water and N-CPHPC within the 2Fo-Fc electron density map (grey mesh,  $1.5\sigma$ ). Clear evidence can be seen in the electron density map for the *trans* arrangement of the peptide bond between the C-terminal D-proline and hGlu. Image generated in Pymol.

One of the key considerations in the design of the multivalent ligands for SAP was the length of the linker used to separate the headgroup from the central cyclic core. The linkers needed to be long enough to allow the headgroups to reach all five binding sites of SAP simultaneously without imposing a strained conformation on the headgroup. The NADPro ligand from the NADPro-SAP complex was superimposed on the SAP-**24** crystal structure to assess the fit of the D-proline headgroups within binding site of SAP (Figure 5.13). The position of NADPro and the D-proline termini of **24** show very close correlation in the position of the carboxyl and pyrrolidine functional groups. However, the orientation of the peptide bond at the N-terminus of D-proline is different. In NADPro-SAP, all acetyl groups of the ligand molecules were built with a *cis* conformation, in the SAP-**24** structure there is clear evidence that the peptide bond adopts the *trans* conformation. This appears to have been due to the geometric constraints imposed on hGlu by the chiral  $\alpha$ -carbon. Therefore, the design of the



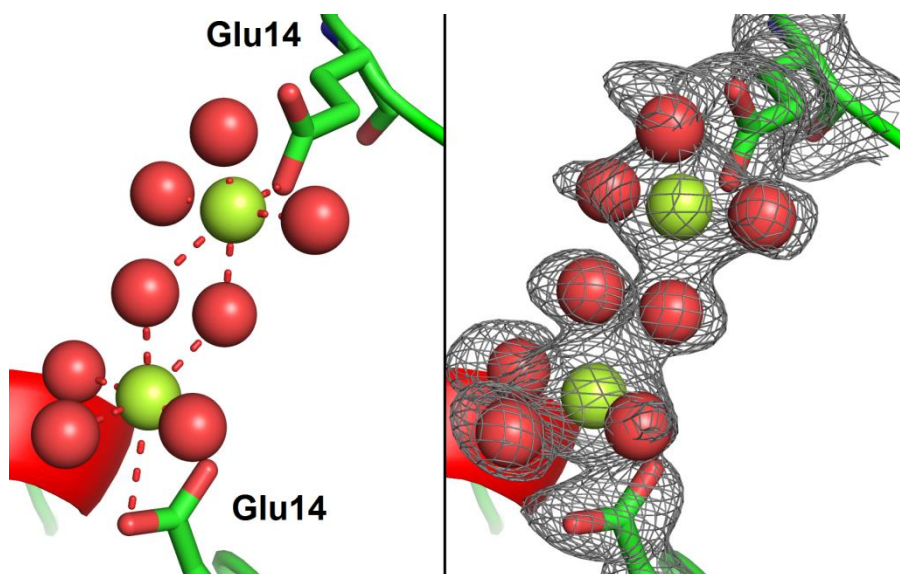
headgroups has not conserved the weak hydrogen bond between the carbonyl oxygen of the peptide bond and the amide side chain of Gln148 identified in Chapter 2. Although this may have been beneficial, the binding affinities measured by ITC suggest that the loss of this bond has not been detrimental to the overall binding mechanism. This emphasises the fact that the coordination of the metal ions and packing of the pyrrolidine ring into the hydrophobic pocket are the key interactions for small molecule binding to SAP and these feature have been retained in ligand **24** despite its more complex structure.



**Figure 5.13.** Comparison of D-proline position for N-CPHPC (turquoise carbons) and NADPro (yellow carbons) when bound in the calcium-binding site of SAP. There are significant similarities in the position of the two D-proline-based ligands. The key difference is in the arrangement of the peptide bond between the N-terminus of D-proline and the next functional group. NADPro favours a *cis* conformation while the position of the alkyl chain in N-CPHPC imposes a *trans* conformation. The pyrrolidine ring of NADPro penetrates slightly deeper into the hydrophobic pocket of SAP, this could be a sign of strain in the N-CPHPC headgroup, which the D-proline from adopting the ideal bonding conformation. Model colours: SAP carbons (green), N-CPHPC carbons (turquoise), NADPro carbons (yellow), oxygen (red), nitrogen (blue), water (red spheres), cadmium ions (limon yellow). Figure generated in Pymol.

The multivalent ligand is not the only species bridging a non-covalent interaction between each pentamer in the SAP-**24** crystal structure. Two cadmium ions coordinated by multiple water molecules bridge the space between carboxyl groups of Glu14 on

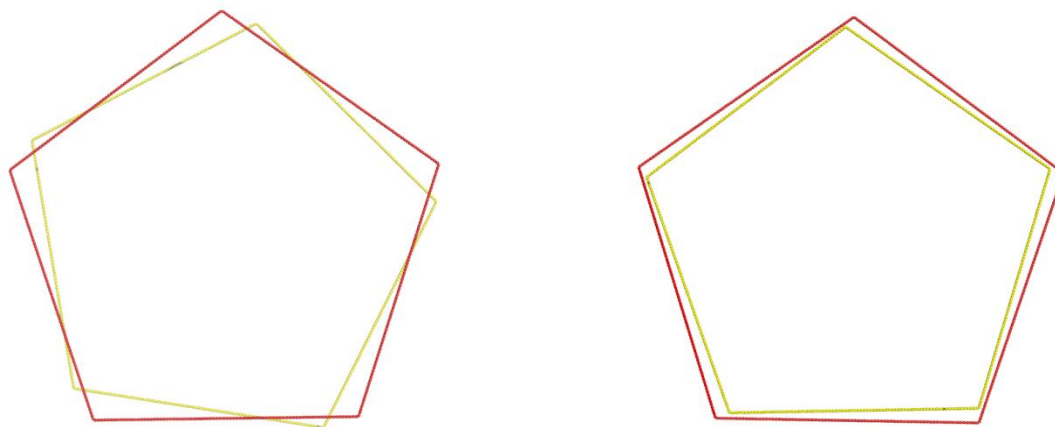
opposing protomers (Figure 5.14). A number of other cadmium sites have been identified with varying degrees of confidence. Coot's unmodelled blob feature highlights difference density in the electron density map that is too large to be a water molecule. Some areas of the electron density map showed roughly spherical density that was determined to be cadmium ions. However, the average B-factor for these ions was 76.72, significantly more than the average value of 33.76 determined for the overall structure by the CCP4 program, Baverage. Therefore, some cadmium ions were built with partial occupancy. It may be that another metal ion was responsible for the peak in the electron density map, such as sodium. However, it has not been possible to definitively identify the ion from the electron density, so the reduced occupancy can be viewed as an expression of doubt in both the position and identity of the ion.



**Figure 5.14.** The electron density map (right panel, grey mesh, 2Fo-Fc map drawn at  $1.5\sigma$ ) shows a number of cadmium ions coordinated by multiple water molecules and the carboxyl side chain of Glu14. The combination of water molecules and cadmium ions appear to form a bridge between Glu14 residues on opposing SAP protomers. Left: red dashed lines show the coordination of cadmium ions (limon yellow) by oxygens from water (red spheres) and carboxyl side chains of Glu14 (red; not to scale). Image generated with Pymol.

Different length alkyl chain linkers in bivalent ligands affect decamer geometry in terms of pentamer separation and relative displacement or “roll” of pentamers when viewed top-down. The pentamer roll for SAP-24 was measured in Pymol to be  $17^\circ$ , which is significantly more than the  $5^\circ$  measured for SAP-CPHPC (Figure 5.15). Additionally,

inspection of Figure 5.15 shows that the roll was anti-clockwise for the uppermost pentamer in SAP-**24** and clockwise for SAP-CPHPC. The pentamers are also 1.6Å closer in SAP-**24** when the distance between identical residues from opposing pentamers is normalised for the increased pentamer roll.



**Figure 5.15.** Roll of pentamers in SAP-**24** and SAP-CPHPC (PDB file: 4AVV; Kolstoe *et al.*, 2014). Distances and pentamer roll measured in Pymol. Pentamer roll determined using position of  $\alpha$ -carbon of residue 53, which sits close to the middle of each protomer.

The smaller pentamer separation and increased pentamer roll is likely to be due to the nature of the ligands bridging the gap between the two pentamers in the decamer. Therefore, the multivalent headgroup may have played some part in the arrangement. The hydrated cadmium complexes may also have contributed to the relative displacement (Figure 5.14). The strong density observed for the alkyl chain between the D-proline termini in the N-CPHPC ligand suggests a defined conformation. This is in contrast to previous reports of SAP-CPHPC where the electron density for the alkyl chain is less clear and the conformation often indeterminable (Kolstoe *et al.*, 2014). By occupying a single defined conformation, the ligand is likely to be forming a more rigid complex with SAP, which may display greater stability because the thermal motion of the complex is less, reducing the probability that the complex will dissociate. The multivalent ligand will have a greater reduction in conformational entropy by adopting a defined conformation, but considering the disorder observed for the rest of the multivalent ligand, it is unlikely that this greatly influenced the stability of the SAP-**24** complex in terms of entropy.

**5.4.6. Cooperative binding of multivalent ligands with SAP**

The multivalent ligands have been shown to participate in cooperative binding with SAP by UV/Vis spectroscopy, isothermal titration calorimetry and x-ray crystallography. Mass spectrometry and x-ray crystallography have provided evidence that decavalent ligands can be synthesised with greatly increased mass and complexity than the bivalent ligand upon which they are based and still maintain the non-covalent cross-linking ability of CPHPC (Pepys *et al.*, 2002). X-ray crystallography of the SAP-**24** complex has shown multivalent ligands to bind to SAP through their carboxyl termini, preserving the binding mechanism of the univalent ligand upon which they are based. Finally, ligands were shown by ITC to have binding affinities almost  $6 \times 10^5$  times stronger than univalent N-acetyl-D-proline. Furthermore, ligands **25** and **26** may have tighter binding affinities for SAP than CPHPC, which is the strongest known small molecule ligand for SAP reported in the literature (Pepys *et al.*, 2002).

There are clear areas for improvement in the synthesis, sample preparation and analysis of these multivalent ligands. In particular, the purity of the multivalent ligands needs to be increased. This can be addressed through minor changes to the structure of the ligands and improved synthesis and purification protocols, which will be discussed in the next chapter. The attained purity of the multivalent ligands has not been sufficient to provide reliable data on binding stoichiometry by ITC, but the results are in agreement with previous ITC measurements of N-acetyl-D-proline and CPHPC recorded by Kolstoe (Kolstoe thesis, 2005). Therefore, conclusions about stoichiometry from the ITC data have still been possible and reinforced by the decameric complex observed in the SAP-**24** crystal structure. Overall these data have provided the first recorded example of synthesis of multivalent ligands for SAP. This is the first time multivalent structures designed for other proteins (Zhang *et al.*, 2004) have been applied to SAP, so it is to be expected that the synthetic strategy and methods of analysis need optimisation and suggestions for doing so will be introduced in the next Chapter.

## Chapter 6. Conclusions

Despite long-standing evidence of the pentraxins homo-pentameric structure suggesting that subunits cooperate in binding multiple ligands simultaneously (Osmand *et al.*, 1977), this work comprises the first reported synthesis and characterisation of multivalent ligands binding simultaneously to all five binding sites of Serum Amyloid P component (SAP). The fact that multivalent scaffolds displaying sub-nanomolar binding affinities for SAP have been constructed using only commercially available building blocks, solid phase peptide synthesis (SPPS) and basic side chain modifications demonstrates the strength of peptide-based architectures.

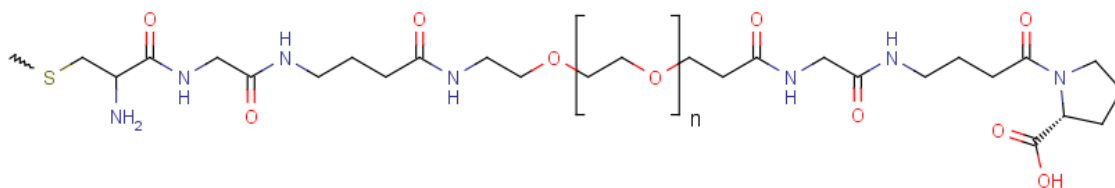
The idea for multivalent ligands targeting the pentraxins was first seriously considered after bivalent SAP ligands were discovered during a search of a Roche compound library (Pepys *et al.*, 2002). The D-enantiomer of Captopril was one of the initial hit compounds and with some serendipity it was discovered that dimerising the ligand by disulphide formation produced a bivalent compound with significantly enhanced binding affinity for SAP. CPHPC was the result of optimising the functional groups of the bivalent D-Captopril hit compound. CPHPC is clearly a very promising therapeutic; it has a simple structure, is not metabolised *in vivo* and binds to SAP with a low nanomolar dissociation constant (Gillmore *et al.*, 2010). However, CPHPC is incapable of removing SAP from the surface of amyloid fibrils (Tennent *et al.*, 1995). Development of an anti-SAP antibody is an alternative strategy that only requires removal of circulating SAP and relies on its persistence in all forms of amyloid (Bodin *et al.*, 2010). Antibodies are of course notoriously expensive to produce (Samaranayake *et al.*, 2009), so finding a molecule that can compete with amyloid deposits for bound SAP offers the potential for a single drug treatment to stimulate removal of all forms of amyloid.

It has not been possible to measure the exact binding affinity with SAP for some of the decavalent ligands synthesised because their projected binding affinity is beyond the measurable limit of isothermal titration calorimetry. Pursuing a method that can accurately measure the concentration of unbound ligand should be the next logical step.

If accurate determination confirms the sub-nanomolar binding affinity recorded in Chapter 5, then there is a realistic probability that the decavalent ligands will be able to competitively inhibit SAP binding to amyloid deposits.

Competitive inhibition experiments would be relatively straightforward. Amyloid fibres could be prepared *in vitro* (Reviewed by Nilsson, 2004). After SAP has been incubated with the fibres, any unbound SAP would be washed from the fibres and a solution containing the decavalent ligand added. After a suitable period of time the multivalent ligand solution could be separated from the fibres and analysed by gel filtration. If SAP has been removed from the fibres then it will be present as a decameric complex producing a peak that can be quantified by using a series of reference SAP-decavalent ligand samples of known concentration.

A number of recommendations can be made to improve the structure of the multivalent ligands for future investigations based on observations made during synthesis and analysis. Diketopiperazine (DKP) formation is an issue with the pentavalent ligands in their current form. DKP is a recognised problem in solid phase peptide synthesis (SPPS; Barany and Merrifield, 1977). Occurrence of DKP is sequence specific, with peptides containing C-terminal proline particularly susceptible. DKP formation requires two  $\alpha$ -amino acids at the C-terminus of the peptide. Therefore, the observed DKP can be entirely eliminated from future syntheses of pentavalent ligands by swapping the penultimate amino acid glycine with its neighbouring  $\gamma$ -aminobutyric acid (Figure 6.1).

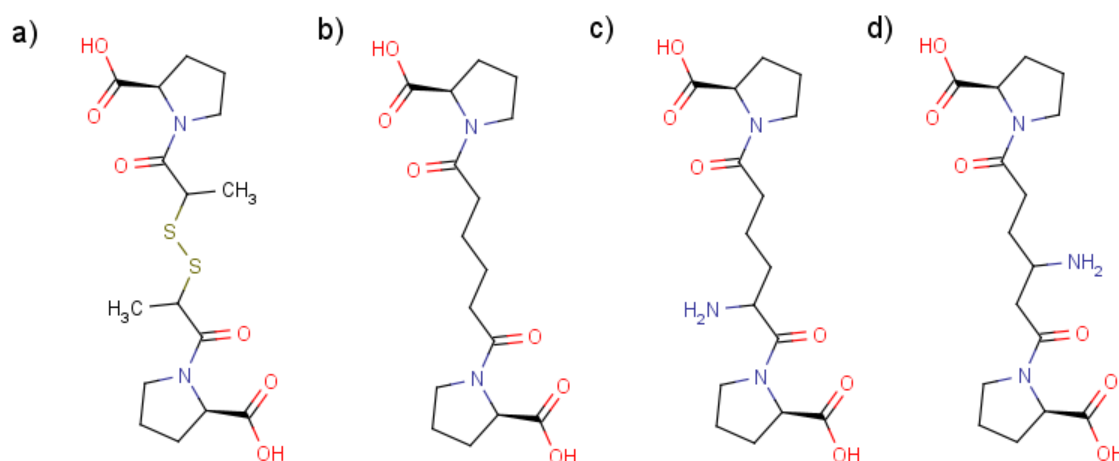


**Figure 6.1.** The proposed structure of linker-headgroups for synthesising pentavalent ligands not susceptible to diketopiperazine formation.

Research of the literature uncovered a potential structural change for the decavalent ligands that may increase the binding affinity by reducing steric hindrance caused by the position of the amino group used to ligate the modified CPHPC headgroup to the multivalent scaffold. Dimerised D-Captopril presented as an SAP ligand by Pepys *et al.*

(2002), showed a much weaker affinity for SAP than CPHPC (Figure 6.2). The removal of a methyl group and a disulphide bond are the structural differences that facilitated tighter binding with SAP. It is not clear whether the improved affinity derives from the loss of the hydrophobic methyl groups, or the fact that CPHPC has greater conformational freedom by eliminating two chiral centres and the disulphide bond. In the decavalent ligands, the headgroup is bound to the multivalent scaffold by replacing the hexanoyl chain of CPHPC with homo-glutamic acid (hGlu). The amino group is attached to the  $\alpha$ -carbon, in a similar position to the methyl groups of dimerised D-Captopril. This could potentially be reducing the binding affinity of the ligands by bringing the attached PEG-linker chain into close contact with the surface of SAP. Using  $\beta$ -homoglutamic acid would move the chain further from the surface of SAP, potentially permitting tighter binding.

**Figure 6.2.** Bivalent ligands for SAP. a) Dimerised D-Captopril. b) CPHPC. c) Amino-CPHPC used as



the headgroup in decavalent ligands **24-26** for SAP, d) proposed structure of headgroup for decameric SAP. Shifting the amino group towards the centre of the molecule could potentially reduce steric hindrance by the attached linker in a multivalent ligand, which may increase the binding affinity.

Application of the multivalent construct presented in this investigation could also be applied to other pentraxins. Multivalent ligands, beyond bivalency, are yet to be considered for C-reactive Protein (CRP), despite an ongoing drug development program based around bis-phosphocholine compounds to target CRP mediated ischemic injury in cardiovascular disease (Pepys *et al.*, 2006). Phosphocholine-based bivalent ligands are promising lead compounds, but they have shown themselves to be difficult to synthesise

in significant quantities at acceptable purity. bis-phosphocholine ligands bind to CRP through the same pharmacological mechanism as CPHPC binding to SAP. The binding of a univalent phosphocholine compound to CRP has a dissociation constant ( $K_d$ ) of  $2\mu\text{M}$  (Kolstoe thesis, 2005). When two phosphocholine residues are linked by a heptanoyl chain the dissociation constant drops to around  $\sim 150\text{nM}$ . In this investigation, linking five N-acetyl D-proline ligands into a single pentavalent compound reduced the dissociation constant from  $15\mu\text{M}$  to around  $40\text{nM}$ . This dramatic increase in binding affinity shows that univalent ligands with only modest binding affinities can be transformed into very strong binding species by combining them into a single multivalent ligand. If this methodology were applied to CRP, it could allow the use of univalent ligands that bind more weakly than phosphocholine, but which has superior qualities in terms of ease of synthesis, purification and sample handling.

The library of small molecules known to bind to CRP is much smaller than for SAP, with phosphocholine and phosphoethanolamine probably the best characterised ligands for CRP. Phosphoethanolamine binds to CRP with a dissociation constant of approximately  $40\mu\text{M}$ . But phosphoethanolamine does not have the positively charged choline group that contributes to difficulties in synthesis of bis-phosphocholine compounds. The synthesis of multivalent phosphoethanolamine compounds is in theory not too complex and could be based on the method used to synthesise bis-phosphocholine complexes (Pepys *et al.*, 2006). Theoretically, linking five phosphoethanolamine compounds into a single pentavalent entity should result in a greatly increased binding affinity, not too dissimilar to that of bis-phosphocholine. And if a decavalent ligand involving bis-phosphoethanolamine headgroups could be synthesised then it would be expected that the affinity would exceed that of the bis-phosphocholine ligands. Furthermore, as the headgroups would make up only a small proportion of the multivalent ligand, their properties should not significantly influence ligand solubility or purifiability.

The obvious objection to using decavalent ligands based on phosphoethanolamine to target CRP is that they would also target SAP. However, evidence of sustained depletion of circulating SAP from amyloidosis patients has shown no ill effect (Gillmore *et al.*,



2010). Only temporary depletion of CRP would be needed to prevent ischemic injury during cardiac events, so the expected inability of bis-phosphoethanolamine compounds to differentiate between human pentraxins would probably not be too detrimental.

This study has also presented some evidence that multivalent ligands can be used to solve difficult crystallographic problems and their use is not limited to studying binding mechanisms or developing therapeutics. Structural biology plays a crucial role in evolving understanding of natural multivalent processes. It is likely that x-ray crystallography will remain a central tool in investigations of multivalent processes and tools that aid the crystallisation processes will always be a benefit to these investigations.

## **7. Acknowledgements**

Thank you Rosie for your constant support and encouragement, I got there in the end. I am sorry I made you move out of London. Hopefully we'll never move away from a capital city again!

I would like to thank the Biotechnology and Biosciences Research Council and Peptide Protein Research Ltd. for funding this project.

Professor Steve Wood and Dr. Robert Broadbridge deserve special mention for their constant advice, encouragement and support (and in Rob's case a roof over my head for a few months too). Thank you Claire and Steve Council for housing me for 9 months, I hope at least it served as a dry run for what to expect when Elliot becomes a teenager. Thanks too to the synthesis and purification teams at PPR, especially when I was under foot.

Thank you to Professor Sir Mark Pepys's laboratory for their generous contribution of SAP and rCRP. Peter Sharratt, Idlir Liko and Talha Arooz provided valuable technical support in conducting amino acid analysis, mass spectrometry and ITC experiments, respectively.

## 8. References

Abernathy T. J., and Avery, O. T. *J. Exp. Med.* **73**(2) (1941), 173-182

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L-W., Kapral, G. J., Grosse-Kuntze, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. *Acta Cryst. D* **66** (2010), 213-221

Aebersold, R., and Mann, M. *Nature*. **13** (2003), 198-207

Afonine P. V., Grosse-Kuntze R. W., Echols N., Headd J. J., Moriarty N. W., Mustyakimov M., Terwilliger T. C., Urzhumtsev A., Zwart P. H. and Adams P. D. *Acta Cryst. D* **68** (2012), 352-367

Agrawal, A., and Volanakis, J. E. *J. Immunol.* **152**(11) (1994), 5404-5410

Anderson, H. L., Anderson, S., and Sanders, J. K. M. *J. Chem. Soc. Perkin. Trans.* **1** (1995), 2231-2245

Andrushchenko, V. V., Vogel, H. J., and Prenner, E. J. *J. Pept. Sci.* **13** (2007), 37-43

Ashton, A. W., Boehm, M. K., Gallimore, J. R., Pepys, M. B., and Perkins, S. J. *J. Mol. Biol.* **272**(3) (1997), 408-422

Atherton, E. *Solid Phase Peptide Synthesis: A Practical Approach*. Sheppard, R. A. (ed.) Oxford University Press. Oxford, UK. 1989

Axford, D., Owen, R. L., Aishima, J., Foadi, J., Morgan, A. W., Robinson, J. I., Nettleship, J. E., Owens, R. J., Moraes, I., Fry, E. E., Grimes, J. M., Harlos, K., Kotecha, A., Ren, J., Sutton, G., Walter, T. S., Stuart, D. I., and Evans, G. *Acta Cryst. D* **68**, (2012), 592-600

## References

Baltz, M. L., De Beer, F. C., Feinstein, A., Mun, E. A., Milstein C. P. Fletcher, T. C., March, J. F., Taylor, J., Bruton, C., Clamp, J. R., Davies, A. J. S., and Pepys, M. B., *Ann. NY Acad. Sci.* **389** (1982), 49-75

Barany, G., and Merrifield, R. B. *J. Am. Chem. Soc.* **99** (1977), 7363

Bickerstaff, M. C. M., Botto, M., Hutchinson, W. L., Herbert, J., Tennent, G. A., Bybee, A., Mitchell, D. A., Cook, H. T., Butler, P. J. G., Walport, M. J., and Pepys, M. B. *Nature Med.* **5(6)** (1999), 694-697

Bijl, M., Horst, G., Bijzet, J., Bootsma, H., Limburg, P. C., and Kallenberg, C. G. M. *Arthritis & Rheumatism.* **48(1)** (2003), 248-254

Bladen, H. A., Nylen, M. U., and Glenner, G. G. J. *Ultrastruct. Res.* **14** (1966), 449-459

Blandamer, M. J., Cullis, P. M., and Engberts, J. B. F. N. *J. Chem. Soc. Faraday Trans.* **94(16)** (1998), 2261-2267

Boas, U., and Heegaard, M. H. *Chem. Soc. Rev.* **33** (2004), 43-63

Bodin, K., Ellmerich, S., Kahan, M. C., Tennent, G. A., Loesch, A., Gilbertson, J. A., Hutchinson, W. L., Mangione, P. P., Gallimore, J. R., Millar, D. J., Minogue, S., Dhillon, A. P., Taylor, G. W., Bradwell, A. R., Petrie, A., Gillmore, J. D., Bellotti, V., Botto, M., Hawkins, P. N., and Pepys, M. B. *Nature.* **468** (2010), 93-97

Botto, M., Hawkins, P. N., Bickerstaff, M. C. M., Herbert, J., Bygrave, A. E., McBride, A., Hutchinson, W. L., Tennent, G. A., Walport, M. J., and Pepys, M. B. *Nature Med.* **3(8)** (1997), 855-859

Breathnach, S. M., Kofler, H., Sepp, N., Ashworth, J., Woodrow, D., Pepys, M. B., and Hintner, H. *J. Exp. Med.* **170** (1989), 1433-1438

## References

- Brönnimann, C., Eikenberry, E. F., Horisberger, R., Hülsen, G., Schmitt, B., Schulze-Bries, C., and Tomizaki, T. *Nucl. Instr. Methods in Phys. Res. A*. **510** (2003), 24-28
- Brünger, A. T. *Nature*. **355** (1992), 472-475
- Burmeister, W. P. *Acta Cryst. D***56** (2000), 328-341
- Butler, P. J. G., Tennent, G. A., and Pepys, M. B. *J. Exp. Med.* **172** (1990), 13-18
- Cairo, C. W., Gestwicki, J. E., Kanai, M., and Kiessling, L. L. *J. Am. Chem. Soc.* **124**(8) (2002), 1615-1619
- Cathcart, E. S., Shirahama, T., and Cohen, A. S. *Biochim. Biophys. Acta*. **147** (1967), 392-393
- Carpino, L. A., and Han, G. Y. *J. Org. Chem.* **37**(22) (1972), 3404-3409
- Chen, V. B., Arendall, III, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. *Acta Cryst. D***66** (2010), 12-21
- Chittasupho, C. *Therapeutic Delivery*. **3**(10) (2012), 1171-1187
- Choi, S.-K. *Synthetic Multivalent Molecules: Concepts and Biomedical Applications*. John Wiley and Sons, Inc. Hoboken, NJ, 2004.
- Cleary, J. P., Walsh, D. M., Hofmeister, J. J., Shankar, G. M., Kuskowski, M. A., Selkoe, D. J., and Ashe, K. H. *Nature Neuromed.* **8**(1) (2005), 79-84
- Coker, A. R., Purvis, A., Baker, D., Pepys, M. B., and Wood, S. P. *FEBS Letts.* **473** (2000), 199-202
- Culpepper, M. A., and Rosenzweig, A. C. *Biochemistry*. **53**(39) (2014), 6211-6219

## References

- Dam, T. K., Roy, R., Das, S. K., Oscarson, S., and Brewer, C. F. *J. Biol. Chem.* **275**(19) (2000), 14223-14230
- Dawson, P. E., Churchill, M. J., Ghadiri, M. R., and Kent, S. B. H. *J. Am. Chem. Soc.* **119**(19) (1997), 4325-4329
- de Beer, F. C., and Pepys, M. B. *J. Immunol. Methods.* **50** (1982), 17-31
- Deniaud, D., Julienne, K., and Gouin, S. G. *Org. Biomol. Chem.* **9** (2011), 966-979
- Dettmar, A. K., Binder, E., Greiner, F. R., Liebau, M. C., Kurschat, C. E., Jungraithmayr, T. C., Saleem, M. A., Schmitt, C-P., Feifel, E., Orth-Höller, D., Kemper, M. J., Pepys, M. B., Würzner, R., and Oh, J. *Infection and Immunity.* **82**(5) (2014), 1872-1879
- Diamond Light Source. *3D model of Diamond showing individual beamlines.* Diamond Light Source. Didcot, UK. (2014). <http://www.diamond.ac.uk/PressOffice/MediaResources.html> accessed on 27/4/2014
- Diamond Light Source. *Beamline Schematic.* Diamond Light Source. Didcot, UK. (2014). <http://www.diamond.ac.uk/Beamlines/Mx/I03/Beamline-Schematic.html> accessed on 27/4/2014
- Diederichs, K., and Karplus, P. A. *Nature Struct. Bio.* **4** (1997), 269-275
- Dodds, D. C., Omeis, I. A., Cushman, S. J., Helms, J. A., and Perin, M. S. *J. Biol. Chem.* **272**(34) (1997), 21488-21494
- Dowton, S. B., and McGrew, S. D. *J. Biochem.* **270** (1990), 553-556
- Dowton, S. B., Waggoner, D. J. *J. Immunol.* **143**(11) (1989), 3776-3780

## References

- Edelman, G. M., Reeke, G. N., Wang, J. L., Waxdal, M. J., Becker, J. W., and Cunningham, B. A. *Proc. Natl. Acad. Sci. USA*. **69(9)** (1972), 2580-2584
- Edman, P., Högfeldt, E., Sillén, L. G., and Kinell, P-O. *Acta Chem. Scand.* **4** (1950), 283-293
- Emsley, J., White, H. E., O'Hara, B. P., Oliva, G., Srinivasan, N., Tickle, I. J., Blundell, T. L., Pepys, M. B., and Wood, S. P. *Nature*. **367** (1994), 338-345
- Emsley, P., and Cowtan, K. *Acta Cryst.* **D60** (2004), 2126-2132
- Ercolani, G. *J. Am. Chem. Soc.* **125(51)** (2003), 16097-16103
- Ercolani, G., and Schiaffino, L. *Angew. Chem. Int. Ed.* **50** (2011), 1762-1768
- Evans, P. *Acta Cryst.* **D62** (2006), 72-82
- Familian, A., Zwart, B., Huisman, H. G., Rensink, I., Roem, D., Hordijk, P. L., Aarden, L. A., and Hack, C. E. *J. Immunol.* **167(2)** (2001), 647-654
- Felluga, F., Tecilla, P., Hillier, L., Hunter, C. A., Licini, G., and Scrimin, P. *Chem. Commun.* **12** (2000), 1087-1088
- Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., Whitehouse, C. M. *Science*. **246** (1989), 64-71
- Ferrige, A. G., Seddon, M. J., and Jarvis, S. *Rapid Comm. in Mass Spec.* **5** (1991), 374-379

## References

- Fields, G. B., Carr, S. A., Marshak, D. R., Smith, A. J., Stults, J. T., Williams, L. C., Williams, K. R., and Young, J. D. *Techniques in Protein Chemistry IV: Evaluation of peptide synthesis as practised in 53 different laboratories*. Hogue-Angeletti, R. (ed.) Academic Press. San Diego, CA. (1993), 229-237
- Freyer, M. W., and Lewis, E. A. *Methods in Cell Biol.* **84** (2008), 79-113
- Friedrich, N., and Kekule, A. *Arch. Pathol. Anat. Physiol. Klin. Med.* **16** (1859), 50-65
- Gabay, C., and Kushner, I. *New Eng. J. Med.* **340(6)** (1999), 448-454
- Gaboriaud, C., Juanhuix, J., Gruez, A., Lacroix, M., Darnault, C., Pignol, D., Verger, D., Fontecilla-Camps, J. C., and Arlaud, G. J. *J. Biol. Chem.* **278(47)** (2003), 46974-46982
- Galli, C., and Mandolini, L. *Eur. J. Org. Chem.* **18** (2000), 3117-3125
- Garman, E. *Acta. Cryst.* **D55** (1999), 1641-1653
- Garman, E. *Acta Cryst.* **D66** (2010), 339-351
- Gertz, M. A., Lacy, M. Q., and Dispenzieri, A. *Kidney Int.* **61** (2002), 1-9
- Gestwicki, J. E., Cairo, C. W., Strong, L. E., Oetjen, K. A., and Kiessling, L. L. *J. Am. Chem. Soc.* **124** (2002), 14922-14933
- Giancola, C., and Pagano, B. *Top. Curr. Chem.* **330** (2013), 211-242
- Gill, R., Kemp, J. A., Sabin, C., and Pepys, M. B. *J. Cereb. Blood Flow Metab.* **24(11)** (2004), 1214-8
- Gillmore, J. D., Tennent, G. A., Hutchinson, W. L., Gallimore, J. R., Lachmann, H. J., Goodman, H. J. B., Offer, M., Millar, D. J., Petrie, A., Hawkins, P. N., and Pepys, M. B. *Brit. J. Haematol.* **148** (2010), 760-767



## References

- Gisin, B. F., and Merrifield, R. B. *J. Am. Chem. Soc.* **94**(9) (1972), 3102-3106
- Glennner, G. G., Keiser, H. R., Bladen., H. A., Cuatrecasas, P., Eanes, E. D., Ram, J. S., Kanfer, J. N., and Delellis, R. A. *J. Histochem. & Cytochem.* **16**(10) (1968), 633-644
- Goldstein, I. J., Hollerman, C. E., and Smith, E. E. *Biochem.* **4**(5) (1965), 876-883
- Gomez-Garcia, M., Benito, J. M., Rodriguez-Lucena, D., Yu, J. X., Chmurski, K., Mellet, C. O., Gallego, R. G., Maestre, A., Defaye, J., Fernandez, J. M. G. *J. Am. Chem. Soc.* **127**(22) (2005), 7970-7971
- Gorevic, P. D., Munoz, P. C., Casey, T. T., Di Raimondo, C. R., Stone, W. J., Prelli, F. C., Rodrigues, M. M., Poulik, M. D., and Frangione, B. *Proc. Natl. Acad. Sci. USA.* **83** (1986), 7908-7912
- Griselli, M., Herbert, J., Hutchinson, W. L., Taylor, K. M., Sohail, M., Krausz, T., and Pepys, M. B. *J. Exp. Med.* **190**(12) (1999), 1733-1739
- Hamazaki, H. *J. Biol. Chem.* **262**(4) (1987), 1456-1460
- Hardy, J. A., and Higgins, G. A. *Science.* **256** (1992), 184-185
- Hawkins, P. N., Tennent, G. A., Woo, P., and Pepys, M. B. *Clin. Exp. Immunol.* **84** (1991), 308-316
- Hawkins, P. N., and Pepys, M. B. *Eur. J. Nucl. Med.* **22** (1995), 595-599
- Headd, J. J., Echols, N., Afonine, P. V., Grosse-Kunstleve, R. W., Chen, V. B., Moriarty, N. W., Richardson, D. C., Richardson, J. S., and Adams, P. D. *Acta Cryst. D* **68** (2012), 381-390
- Hernández, H. and Robinson, C. V., *Nat. Prot.* **2**(3) (2007), 715-726

## References

- Hind, C. R. K., Collins, P. M., and Pepys, M. B. *Biochim. Biophys. Acta.* **802** (1984), 148-150
- Ho, C. S., Lam, C. W. K., Chan, M. H. M., Cheung, R. C. K., Law, L. K., Lit, L. C. W., Suen, M. W. M., and Tai, H. L. *Clin. Biochem. Rev.* **24** (2003), 3-12
- Ho, J. G. S., Kitov, P. I., Paszkiewicz, E., Sadowska, J., Bundle, D. R., and Ng, K. K.-S. *J. Biol. Chem.* **280(36)** (2005), 31999-32008
- Hohenester, E., Hutchinson, W. L., Pepys, M. B., and Wood S. P. *J. Mol. Biol.* **269** (1997), 570-578
- Holmgren, J., and Svennerholm, A. M. *Gastroenterology Clinics of North America.* **21(2)** (1992), 283-302
- Hopkins, M. F., Flanagan, P. A., Bailey, S., Glover, I. D., Myles, D. A. A., and Greenhough, T. J. *J. Mol. Biol.* **235** (1994), 767-771
- Housecroft, C. E., and Constable, E. C. *Chemistry 3<sup>rd</sup> Edition*. Pearson. Edinburgh, UK. (2006)
- Iscove, N. N., Sieber, F., and Winterha, K. H. *J. Cellular. Physiol.* **83(2)** (1974), 309-320
- Isobe, H., Mashima, H., Yorimitsu, H., and Nakamura, E. *Org. Letts.* **5(23)** (2003), 4461-4463
- Jacobsen, H., and Stockmayer, W. H. *Chem. Phys.* **18** (1950), 1600-1606
- Jeffrey, G. A. *An Introduction to Hydrogen Bonding*. Oxford University Press. New York, NY. (1997)

## References

- Jenvey, M. C. Structure Led Drug Design for the Pentraxins. Ph.D Thesis. University of Southampton, UK. (2006)
- Jiménez, J. L., Nettleton, E. J., Bouchard, M., Robinson, C. V., Dobson, C. M., and Saibil, H. R. *Proc. Natl. Acad. Sci. USA*. **99**(14) (2002), 9196-9201
- Jones, D. S., Gamino, M. E., Victoria, E. J., Yu, L., and Coutts S. M. *Tetrahedron Letts.* **39** (1998), 6107-6110
- Jones, D. S., Coutts, S. M., Gamino, C. A., Iverson, G. M., Linnik, M. D., Randow, M. E., Ton-Nu, H-T., and Victoria, E. J. *Bioconjugate Chem.* **10** (1999), 480-488
- Kabsch, W. *XDS. Acta Cryst.* **D66** (2010), 125-132
- Kabsch, W. *J. Appl. Cryst.* **21** (1988), 67-71
- Kanai, M., Mortell, K. H., and Kiessling, L. L. *J. Am. Chem. Soc.* **119** (1997), 9931-9932
- Kane, R. S. *Langmuir.* **26**(11) (2010), 8636-8640
- Kaplan, M. H., and Volanakis, J. E. *J. Immunol.* **112**(6) (1974), 2135-2147
- Kinoshita, C. M., Gewurz, A. T., Siegel, J. N., Ying, S-C., Hugli, T. E., Coe, J. E., Gupta, R. K., Huckman, R., and Gewurz, H. *Protein Sci.* **1** (1992), 700-709
- Kirby, J. A., *Adv. Phys. Org. Chem.* **17** (1980), 183-278
- Kitov, P. I., Sadowska, J. M., Mulvey, G., Armstrong, G. D., Ling, H., Pannu, N. S., Read, R. J., and Bundle, D. R. *Nature.* **403** (2000), 669-672
- Knowles, T. P. J., Vendruscolo, M., and Dobson, C. M. *Nature Rev.* **15** (2014), 384-396

## References

- Kolstoe, S. E. Ligand Binding to Pentraxins. Ph.D Thesis. University of Southampton, UK. (2005)
- Kolstoe, S. E., and Wood, S. P. *Biochem. Soc. Trans.* **38** (2010), 466-470
- Kolstoe, S. E., Jenvey, M. C., Purvis, A., Light, M. E., Thompson, D., Hughes, P., Pepys M. B., and Wood, S. P. *Acta. Cryst.* **D70** (2014), 2232-2240
- Kopský, V., and Litvin, D. B. *International Tables for Crystallography, 2<sup>nd</sup> Edition*. Wiley and Sons. New York, New York. 2010
- Kramer, R. H., and Karpen, J. W. *Nature*. **395** (1998), 710-713
- Krishnamurthy, V. M., Bohall, B. R., Semetey, V., Whitesides, G. M. *J. Am. Chem. Soc.* **128(17)** (2006), 5802-5812
- Krishnamurthy, V. M., Semetey, V., Bracher, P. J., Shen, N., and Whitesides, G. M. *J. Am. Chem. Soc.* **129(5)** (2007), 1312-1320
- Kuhn, W. *Kolloid Zh.* **68** (1934), 2-15
- Lee, G. W., Lee, T. H., and Vilcek, J. *J. Immunol.* **150(5)** (1993), 1804-1812
- Lei, K. J., Liu, T., Zon, G., Soravia, E., Liu, T. Y., and Goldman, N. D. *J. Biol. Chem.* **260** (1985), 13377-13383
- Ling, H., Boodhoo, A., Hazes, B., Cummings, M. D., Armstrong, G. D., Brunton, J. L., and Read, R. J. *Biochemistry*. **37(7)** (1998), 1777-1788
- Lührs, T., Ritter, C., Adrian, M., Riek-Loher, D., Bohrmann, B., Döbeli, H., Schubert, D., and Riek, R. *Proc. Natl. Acad. Sci. USA*. **102(48)** (2005), 17342-17347

## References

- Mammen, M., Choi, S.-K., and Whitesides, G. M. *Angew. Chem. Int. Ed.* **37** (1998a), 2754-2794
- Mammen, M. Shakhnovich, E. I., and Whitesides, G. M., *J. Org. Chem.* **63** (1998b), 3168-3175
- Mantovani, A., Garlanda, C., and Bottazzi, B. *Vaccine*. **21** (2003), S43-S47
- MarvinSketch, version 5.12.0, Chemaxon Ltd. Budapest, Hungary
- Matthews, B. W. *J. Mol. Biol.* **33** (1968), 491-497
- Matsuura, K. *Trends in Glycosci and Glycotech.* **25(144)** (2013), 227-239
- McCoy, A. J. Phase Diagrams. *University of Cambridge*. Cambridge, UK. (2005).  
[http://www-structmed.cimr.cam.ac.uk/Course/Crystals/Theory/phase\\_diag\\_zones.html](http://www-structmed.cimr.cam.ac.uk/Course/Crystals/Theory/phase_diag_zones.html)  
accessed on 9/5/2014
- McCoy A.J., Grosse-Kunstleve R. W., Adams P. D., Winn M. D., Storoni L. C., and Read R. J. *J Appl. Cryst.* **40** (2007), 658-674
- McPherson, A. *Preparation and analysis of protein crystals*. Wiley. New York, NY. (1982)
- McPherson, A. *Eur. J. Biochem.* **189** (1990), 1-23
- Meents, A., Dittrich, B., and Gutmann, S. *J. Synchrotron Rad.* **16** (2009), 183-190
- Meents. A., Gutmann, S., Wagner, A., and Schulze-Briesse, C. *Proc. Natl. Sci. USA.* **107** (2010), 1094-1099
- Merrifield, R. B. *J. Am. Chem. Soc.* **85(14)** (1963), 2149-2154

## References

*MicroCal ITC 200 User Manual*. GE Life Sciences. Uppsala, Sweden. (2012)

Mighell, A. D. *J. Res. Natl. Inst. Stand. Technol.* **107** (2002), 373-377

Mikolajek, S. E. Ligand Binding to Pentraxins. Ph.D Thesis. University of Southampton, UK. (2008)

Mikolajek, H., Kolstoe, S. E., Pye, V. E., Mangione, P., Pepys, M. B., and Wood, S. P. *J. Mol. Recognit.* **24** (2011), 371-377

Milla, P., Dosio, F., and Cattell, L. *Curr. Drug Metabolism.* **13** (2012), 105-119

Mulder, A., Huskens, J., and Reinhoudt, D. N. *Org. Biomol. Chem.* **2** (2004), 3409-3424

Müller, R., Weckert, E., Zellner, J., and Drakopoulos, M. *J. Synchrotron Rad.* **9** (2002), 368-374

Murshudov, G. N., Vagin, A. A., and Dodson, E. J. *Acta Cryst.* **D53** (1997), 240-255

Nelson, S. R., Lyon, M., Gallagher, J. T., Johnson, E. A., and Pepys, M. B. *Biochem. J.* **275** (1991), 67-73

Niederhafner, P., Šebestík, J., and Ježek, J. *J. Pept. Sci.* **14** (2008), 2-43

Nilsson, M. R. *Methods.* **34(1)** (2004), 151-160

Noursadeghi, M., Bickerstaff, M. C. M., Gallimore, J. R., Herbert, J., Cohen, J., and Pepys, M. B. *Proc. Natl. Acad. Sci. USA.* **97(26)** (2000), 14584-14589

O'Brien, R., Chowdry, B. Z., and Ladbury, J. E (eds). *A Practical Approach to Protein-Ligand Interactions: Isothermal Titration Calorimetry*. Oxford University Press. Oxford, UK. 2001, 263-286

## References

- Osmand, A. P., Friedenson, B., Gewurz, H., Painter, R. H., Hofmann, T., and Shelton, E. *Proc. Natl. Acad. Sci USA*. **74**(2) (1977), 739-743
- Padilla, N. D., Bleeker, W. K., Lubbers, Y., Rigter, G. M., Van Mierlo, G. J., Daha, M. R., and Hack, C. E. *Immunol*. **109** (2003), 564-571
- Peisajovich, A., Marnell, L., Mold, C., and Du Clos, T. W. *Expert. Rev. Clin. Immunol*. **4**(3) (2008), 379-390
- Pepys, M. B., Dash, A. C., Munn, E. A. Feinstein, A., Skinner, M., Cohen, A. S., Gewurz, H., Osmand, A. P., and Painter, R. H. *Lancet*. **309** (1977), 1029-1031
- Pepys, M. B., Baltz, M., Gomer, K., Davies, A. J. S., and Doenhoff, M. *Nature*. **278** (1979), 259-261
- Pepys, M. B., and Baltz, M. L. *Adv. Immunol*. **34** (1983), 141-212
- Pepys, M. B., and Butler, P. J. G. *Biochem Biophys. Res*. **148** (1987), 308-313
- Pepys, M. B., Booth, D. R., Hutchinson, W. L., Gallimore, J. R., Collins, P. M., and Hohenester, E. *Int. J. Exp. Clin. Invest*. **4** (1997), 274-295
- Page, R., Grzechnik, S. K., Canaves, J. M., Spraggon, G., Kreusch, A., Kuhn, P., Stevens, R. C., and Lesley, S. A. *Acta. Cryst. D* **59** (2003), 1023-1037
- Pepys, M. B., and Hirschfield, G. M. *J. Clin. Invest*. **111**(12) (2003), 1805-1812
- Pepys, M. B. *Annu. Rev. Med*. **57** (2006), 223-241

## References

- Pepys, M. B., Herbert, J., Hutchinson, W. L., Tennent, G. A., Lachmann, H. J., Gallimore, J. R., Lovat, L. B., Bartfai, T., Alanine, A., Hertel, C., Hoffmann, T., Jakob-Roetne, R., Norcross, R. D., Kemp, J. A., Yamamura, K., Suzuki, M., Taylor, G. W., Murray, S., Thompson, D., Purvis, A., Kolstoe, S., Wood, S. P., and Hawkins, P. N. *Nature*. **417** (2002), 254-259
- Pepys, M. B., Hirschfield, G. M., Tennent, G. A., Gallimore, J. R., Kahan, M. C., Bellotti, V., Hawkins, P. N., Myers, R. M., Smith, M. D., Polara, A., Cobb, A. J. A., Ley, S. V., Aquilina, J. A., Robinson, C. V., Sharif, I., Gray, G. A., Sabin, C. A., Jenvey, M. C., Kolstoe, S. E., Thompson, D., and Wood, S. P. *Nature*. **440** (2006), 1217-1221
- Pierce, M. M., Raman, C. S., and Nall, B. T. *Methods*. **19** (1999), 213-221
- Podolsky, D. K., Weiser, M. M., Lamont, J. T., and Isselbac, K. J. *Proc. Natl. Acad. Sci. USA*. **71**(3) (1974), 904-908
- Pontet, M., Engler, R., and Jayle, M. F. *FEBS Letts*. **88**(2) (1978), 172-175
- Powell, H. R. *Acta Cryst*. **D55** (1999), 1690-1695
- Protein Data Bank. [www.rcsb.org/pdb](http://www.rcsb.org/pdb)
- Puchtler, H., Sweat, F., and Levine, M. *J. Histochem and Cytochem*. **10**(3) (1962), 355
- Purvis, A. Amyloid Recognition by Serum Amyloid P Component. Ph.D Thesis University of Southampton, UK. (2002)
- Pymol. Pymol molecular graphics system, version 1.3. Distributed by Schrödinger LLC
- Qu, B., Li, X., Guan, M., Li, X., Hai, L., and Wu, Y. *Eur. J. Med. Chem*. **72** (2014), 110-118



## References

- Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. *J. Mol. Biol.* **7** (1963), 95-99
- Rao, J., Lahiri, J., Isaacs L., Weis, R. M., and Whitesides, G. M. *Science*. **280** (1998), 708-711
- Rassouli, M., Sambasivam, H., Azadi, P., Dell, A., Morris, H. R., Nagpurkar, A., Mookerjea, S., and Murray, R. K. *J. Biol. Chem.* **267**(5) (1992), 2947-2954
- Ravelli R. B. G., and McSweeney, S. *Structure*. **8** (2000), 315-328
- Ravelli, R. B. G., Theveneau, P., McSweeney, S. and Caffrey, M. *J. Synchrotron Rad.* **9** (2002), 355-360
- Ravelli, R. B. G., and Garman, E. F. *Curr. Opinion. Struct. Bio.* **16** (2006), 624-629
- Read, R. J., *Acta Cryst.* **A42** (1986), 140-149
- Reid, M. S., and Blobel, C. P. *J. Biol. Chem.* **269**(51) (1994), 32615-32620
- Renaudet, O., and Dumy, P. *Org. Letts.* **5**(3) (2003), 243-246
- Rupp, B. *Biomolecular Crystallography*. Garland Science. New York, NY. (2010)
- Sambasivam, H., Rassouli, M., Murray, R. K., Nagpurkar, A., Mookerjea, S., Azadi, P., Dell, A., and Morris, H. R. *J. Biol. Chem.* **268**(14) (1993), 10007-10016
- Samaranayake, H., Wirth, T., Schenkwein, D., Rätty, J. K., and Yiä-Herttuala, S. *Ann. Med.* **41**(5) (2009), 322-331
- Saraiva, M. J. M. *Human Mutation*. **5** (1995), 191-196
- Schön, A., and Freire, E. *Biochem.* **28**(12) (1989), 5019-5024

## References

- Schüttelkopf, A. W., and van Aalten, D. M. F. *Acta Cryst.* **D60** (2004), 1355-1363
- Shrive, A. K., Holden, D., Myles, D. A., and Greenhough, T. J. *Acta Cryst.* **D52** (1996), 1049-1057
- Siegel, J. Rent, R., Gewurz, H. *J. Exp. Med.* **140(3)** (1974), 631-647
- Siegel, R. A., and Linstad, J. L. *J. Phys. Chem. B.* **114** (2010), 14071-14076
- Sigurskjold, B. W., *Anal. Biochem.* **277** (2000), 260-266
- Sipe, J. D. *Annu. Rev. Biochem.* **61** (1992), 947-975
- Smith, C. G., and Vane, J. R. *FASEB.* **17(8)** (2003), 788-789
- Spackman, D. H., Stein, W. H., and Moore, S. *Anal. Chem.* **30(7)** (1958), 1190-1206
- Sorensen, G. D., and Shimamura, T. *Lab. Invest.* **13(11)** (1964), 1409-1417
- Srinivasan, N., White, H. E., Emsley, J., Wood, S. P., Pepys, M. B., and Blundell, T. L. *Structure.* **2** (1994) 1017-1027
- Sunde, M., Serpell, L. C., Bartlam, M., Fraser, P. E., Pepys, M. B., and Blake, C. C. F. *J. Mol. Biol.* **273** (1997), 729-739
- Tam, J. P. *Proc. Natl. Acad. Sci. USA.* **85** (1988), 5409-5413
- Tan, S. Y., and Pepys, M. B. *Histopathology.* **25(5)** (1994), 403-414
- Teng, T., and Moffat, K. *J. Synchrotron Rad.* **7** (2000), 313-317
- Tennent, G. A., and Pepys, M. B. *Biochem. Soc. Trans.* **22** (1994) 74-79

## References

- Tennent, G. A., Lovat, L. B., and Pepys, M. B. *Proc. Natl. Acad. Sci.* **92** (1995), 4299-4303
- Thompson, A. R., and Enfield, D. L. *Biochem.* **17(20)** (1978), 4304-4311
- Thompson, D., Pepys, M. B., Tickle, I., and Wood, S. P. *J. Mol. Biol.* **320** (2002), 1081-1086
- Thompson, D., Pepys, M. B., and Wood, S. P. *Structure.* **7(2)** (1999), 169-177
- Tomalia, D. A., Baker, H., Dewald, J., Hall, M., Kallos, G., Martin, S., Roeck, J., Ryder, J., and Smith, P. *Polymer.* **17(1)** (1985), 117-132
- Trakhanov, S., Kreimer, D. I., Parkin, S., Ames, G. F-L., and Rupp, B. *Protein Sci.* **7** (1998), 600-604
- van Dongen, W. D., and Heck, A. J. R. *Analyst.* **125** (2000), 583-589
- Verkerk, U. H., and Kebarle, P., *J. Am. Chem. Soc. Mass. Spec.* **16(8)** (2005), 1325-1341
- Virchow, R. *Verh. Phys. Med. Ges. Wurzburg.* **2** (1851), 51-54
- Voet, D. *Biochemistry.* 3<sup>rd</sup> Edition. John Wiley and Sons. New York, NY. (1995)
- Volanakis, J. E., and Kaplan, M. H. *Proc. Soc. Exp. Bio. Med.* **136(2)** (1971), 612-614
- Volanakis, J. E., and Wirtz, K. W. A. *Nature.* **281** (1979), 155-157
- Vrasidas, I., Andre, S., Valentini, P., Bock, C., Lensch, M., Kaltner, H., Liskamp, R. M. J., Gabius, H. J., and Pieters, R. J. *Org & Biomol. Chem.* **1(5)** (2003), 803-810

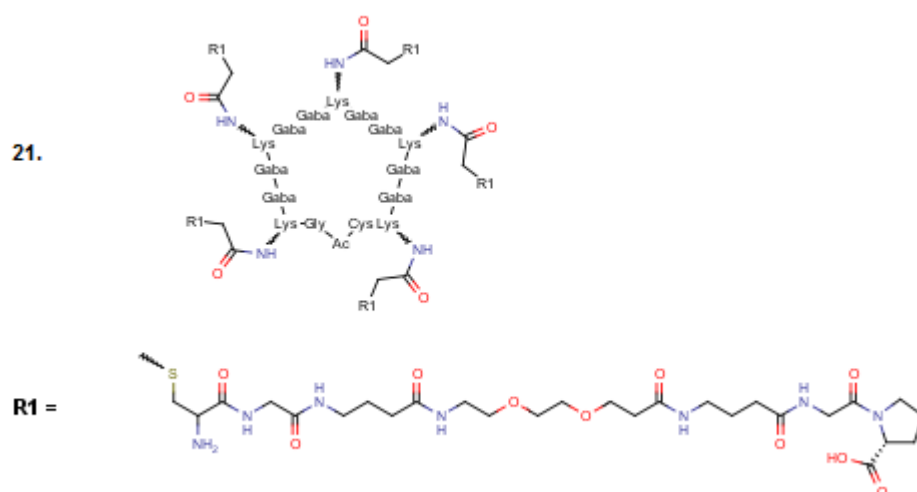
## References

- Wang, Y., Guo, Y., Wang, X., Huang, J., Shang, J., and Sun, S. *Gene Therapy*. **19** (2012), 70-77
- Weik, M., Ravelli, R. B. G., Kryger, G., McSweeney, S., Raves, M. L., Harel, M., Gros, P., Silman, I., Kroon, J. and Sussman, J. L. *Proc. Natl. Acad. Sci. USA*. **97** (2000), 623-628
- Weiss, M. S., and Hilgenfeld, R. *J. Appl. Cryst.* **30** (1997), 203-205
- Werle, M., and Bernkop-Schnürch, A. *Amino Acids*. **30(4)** (2006), 351-367
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. *Acta. Cryst.* **D67** (2011), 235-242
- Winn, M. D., Isupov, M. N., and Murshudov, G. N. *Acta Cryst.* **D57** (2001), 122-133
- Wisemann, T., Williston, S., Brandts, J. F., and Lin, L-N., *Anal. Biochem.* **179** (1989), 131-137
- Woo, P., Korenberg, J. R., and Whitehead, A. S. *J. Biol. Chem.* **260** (1985), 13384-13388
- Yang, G. C. H., Nieto, R., Strachura, I., and Gallo, G. R. *Am. J. Pathol.* **141** (1992), 409-419
- Young, B., Gleeson, M. Cripps, A. W. *Pathology*. **23(2)** (1991), 118-124
- Zhang, Z., Liu, J., Verlinde, C. L. M. J., Hol., W. G. J., and Fan, E. *J. Org. Chem.* **69** (2004), 7737-7740

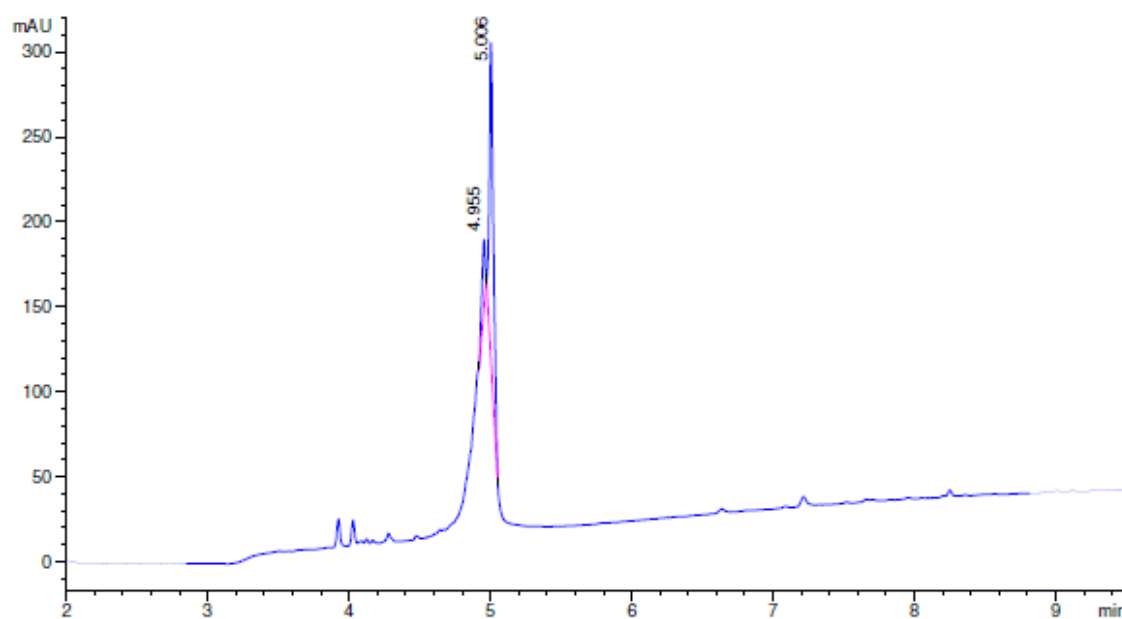
## Appendix

### A.1. Liquid chromatography mass spectrometry analysis of purified multivalent ligands 21-26

#### A.1.1. Compound 21



**Figure A.1.** Structure of compound **21**.



**Figure A.2.** UV-Visible spectrum after liquid chromatography of compound **21** at 280nm.

Peak Number	Retention Time (mins)	Area (%)
1	4.96	14.83
2	5.01	85.17

**Table A.1.** Purity calculation from peak area of LC trace for purified compound **21**.

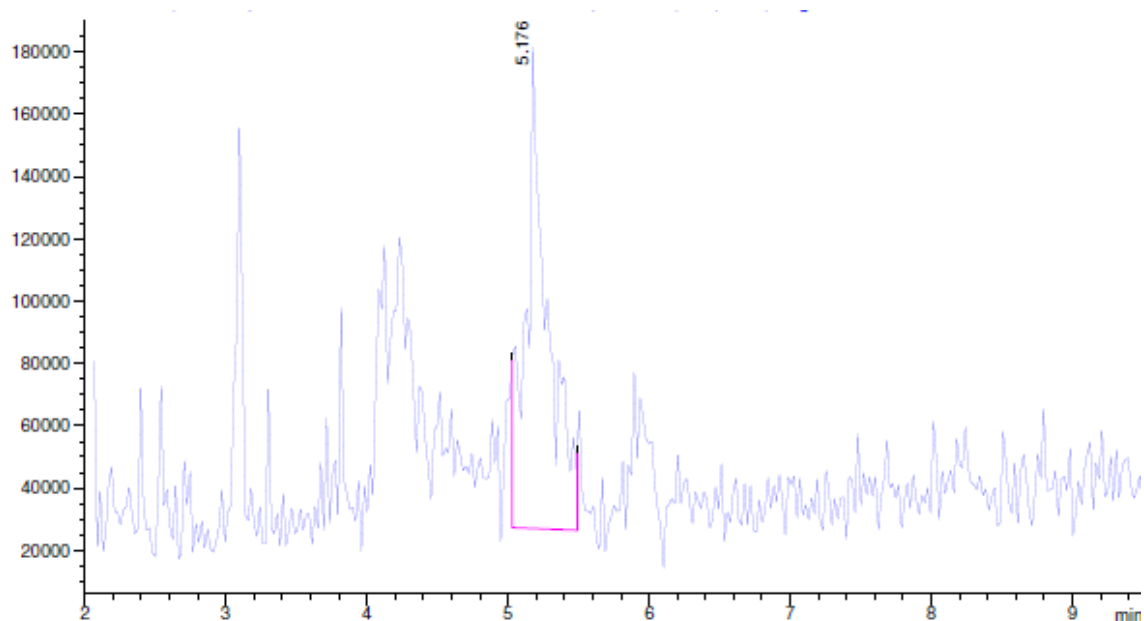


Figure A.3. Mass spectrometry detector total ion count recorded during LCMS of purified **21**.

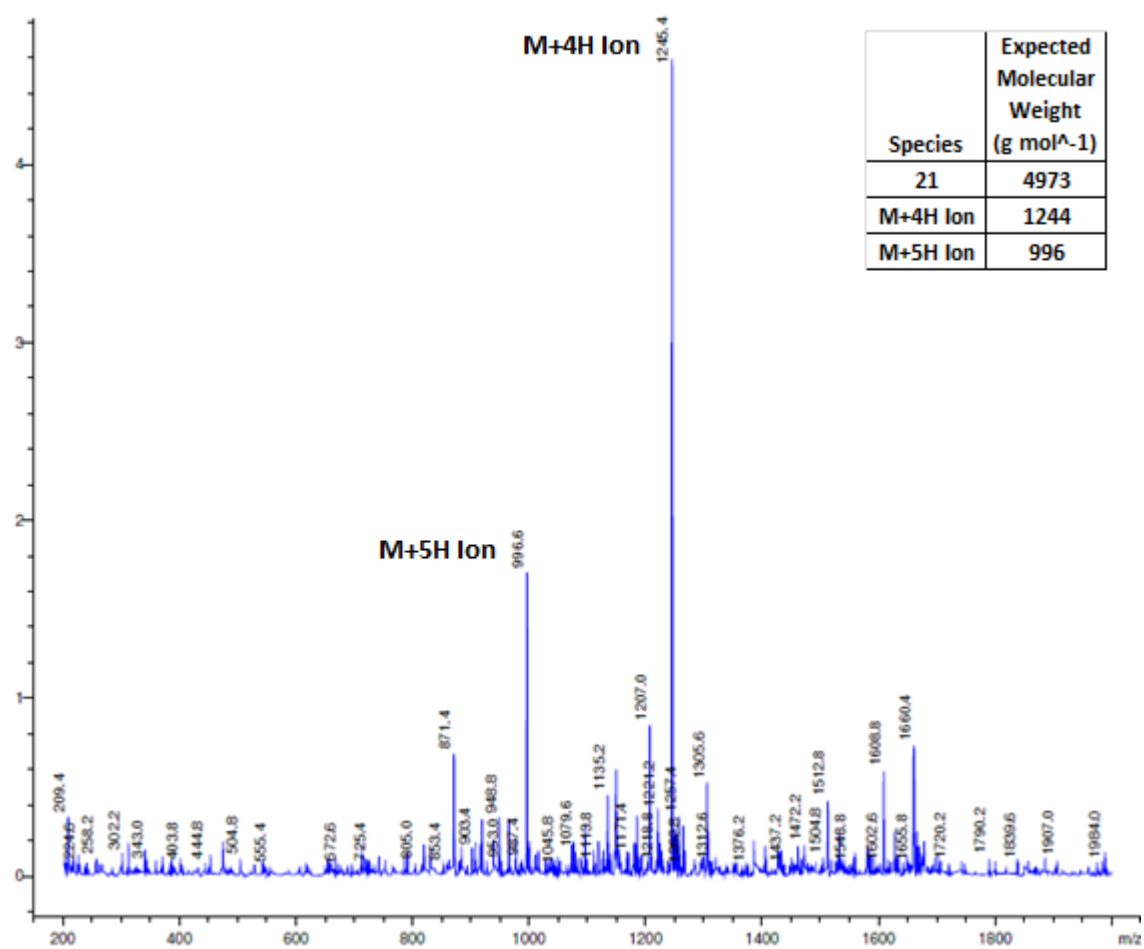


Figure A.4. Mass spectrum of peak detected at 5.18 mins.

## A.1.2. Compound 22

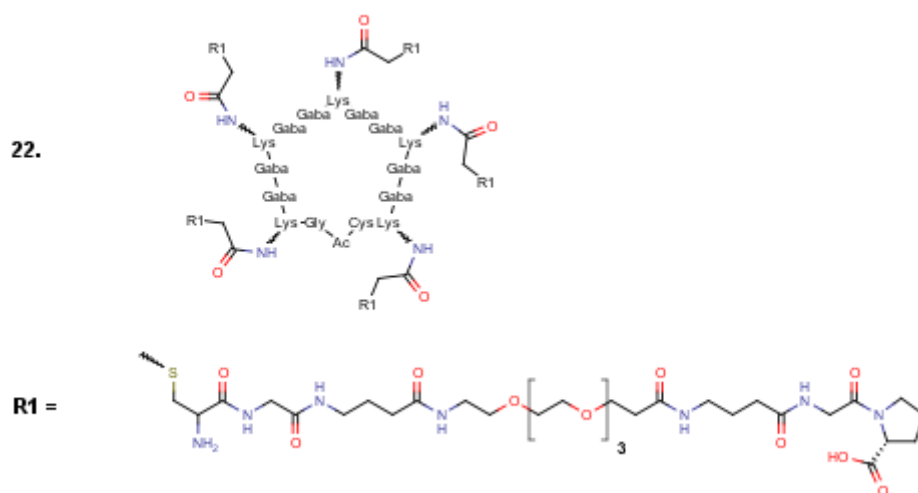


Figure A.5. Structure of compound 22.

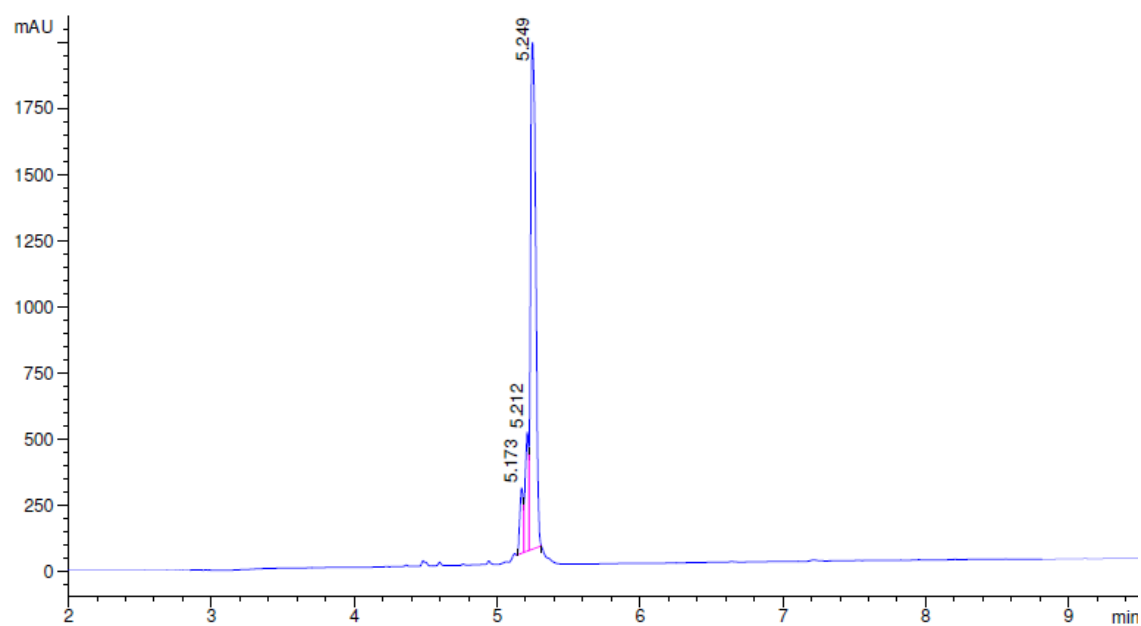
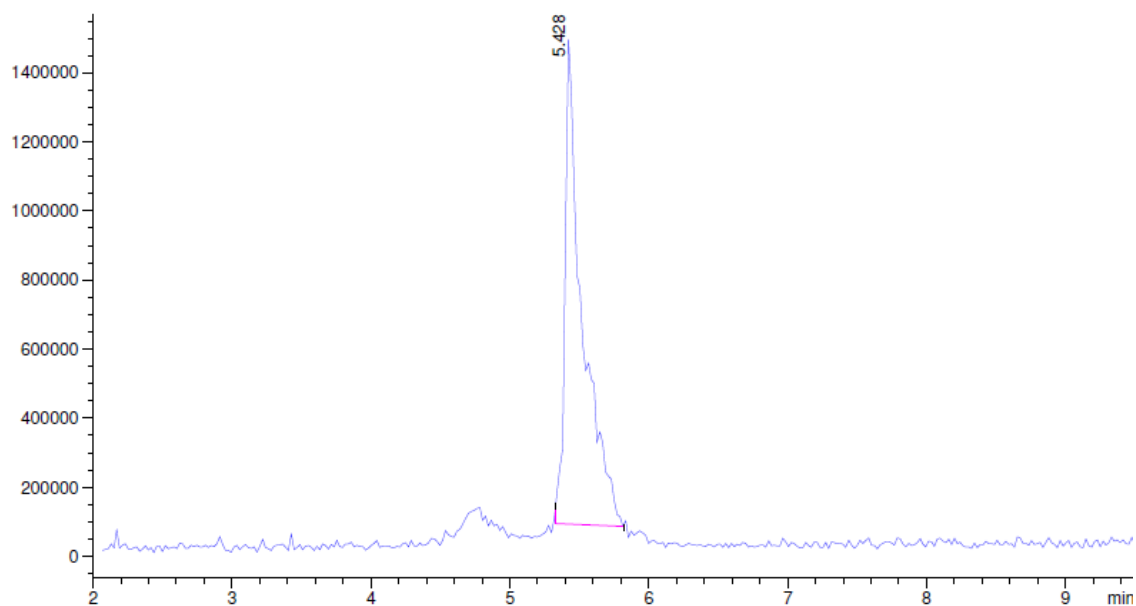


Figure A.6. UV-Vis spectrum after liquid chromatography of compound 22.

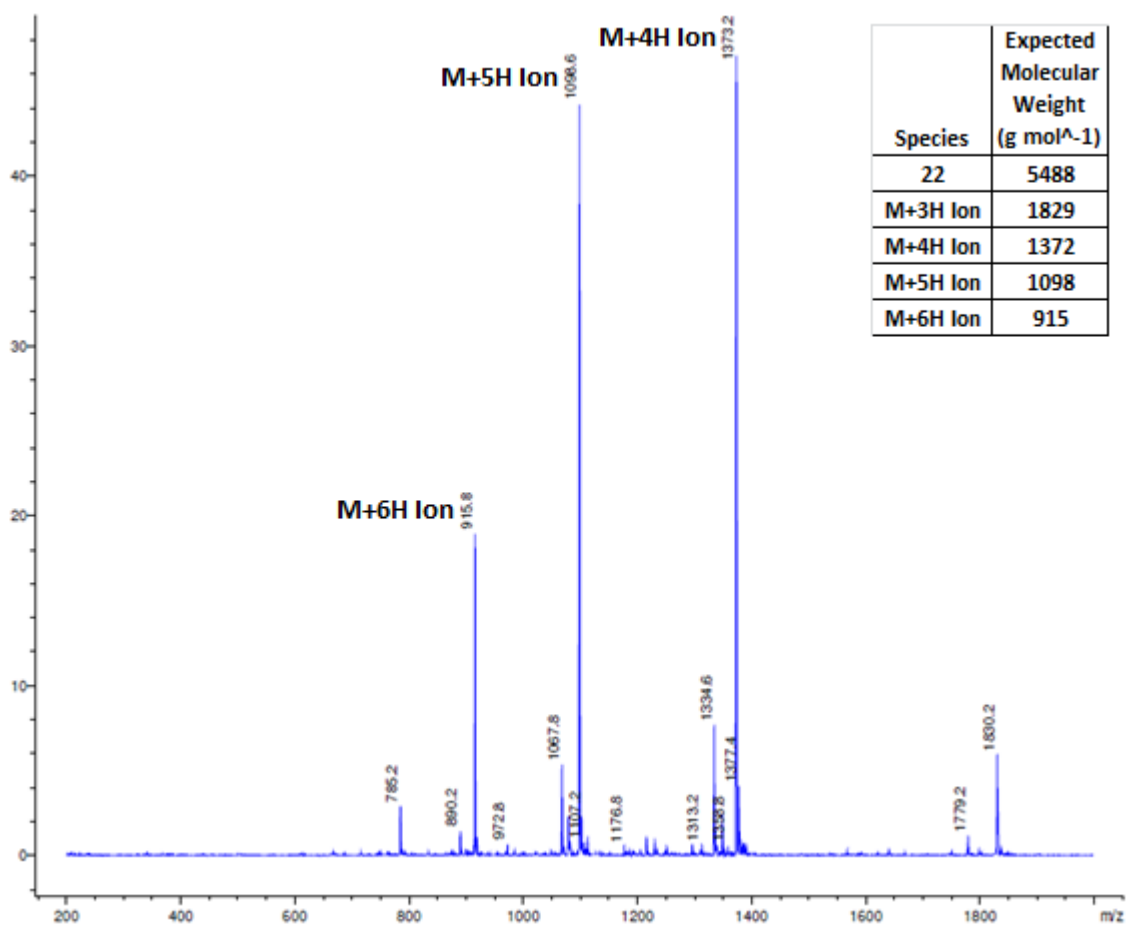
Peak Number	Retention Time (mins)	Area (%)
1	5.17	6.71
2	5.21	12.03
3	5.25	81.27

Table A.2. Purity calculation from peak area of LC trace for purified compound 22.

## Appendix



**Figure A.7.** Mass spectrometry detector total ion count recorded during LCMS of purified **22**.



**Figure A.8.** Mass spectrum of peak detected at 5.43 mins.



## A.1.3. Compound 23

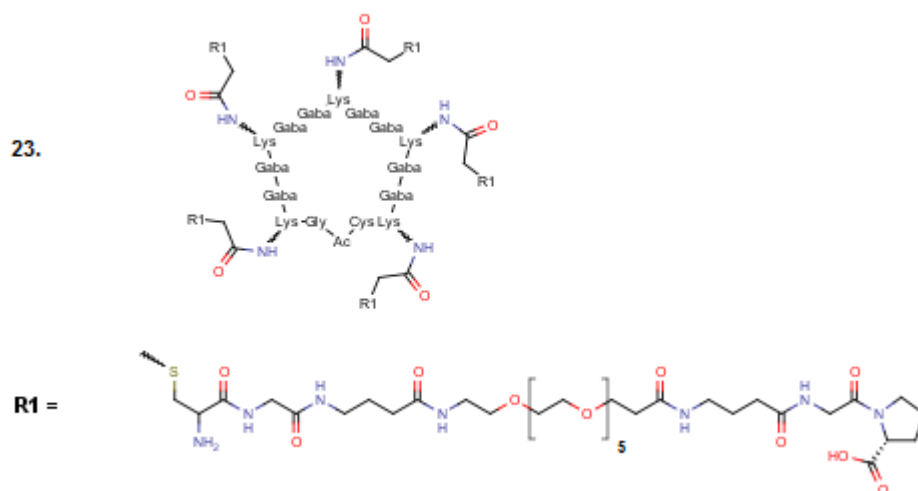


Figure A.9. Structure of compound 23.

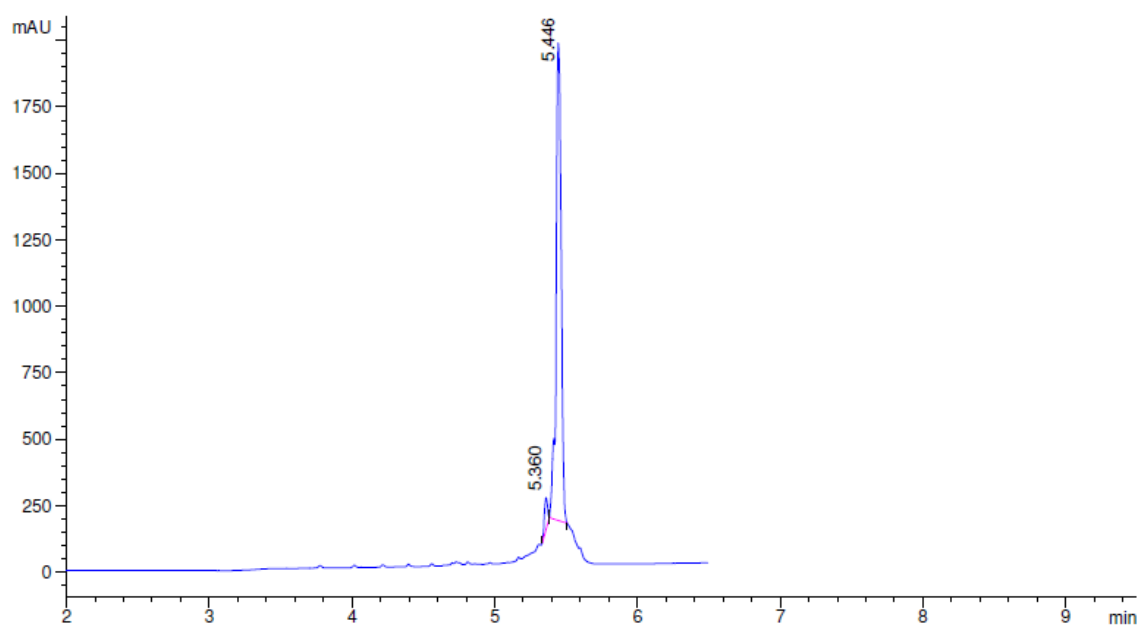
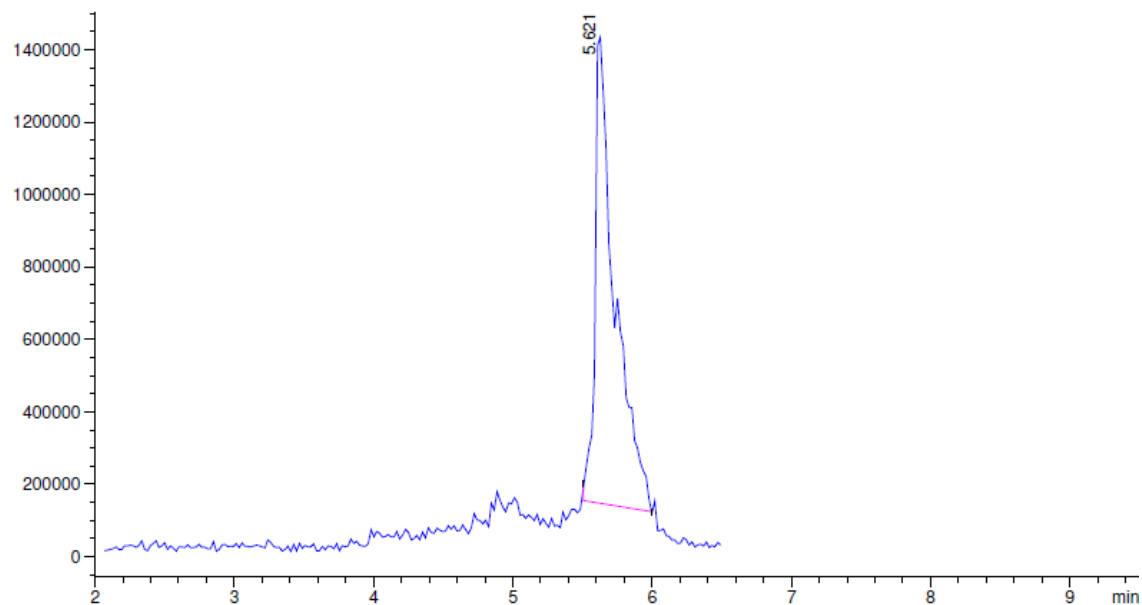


Figure A.10. UV-Vis spectrum after liquid chromatography of compound 23.

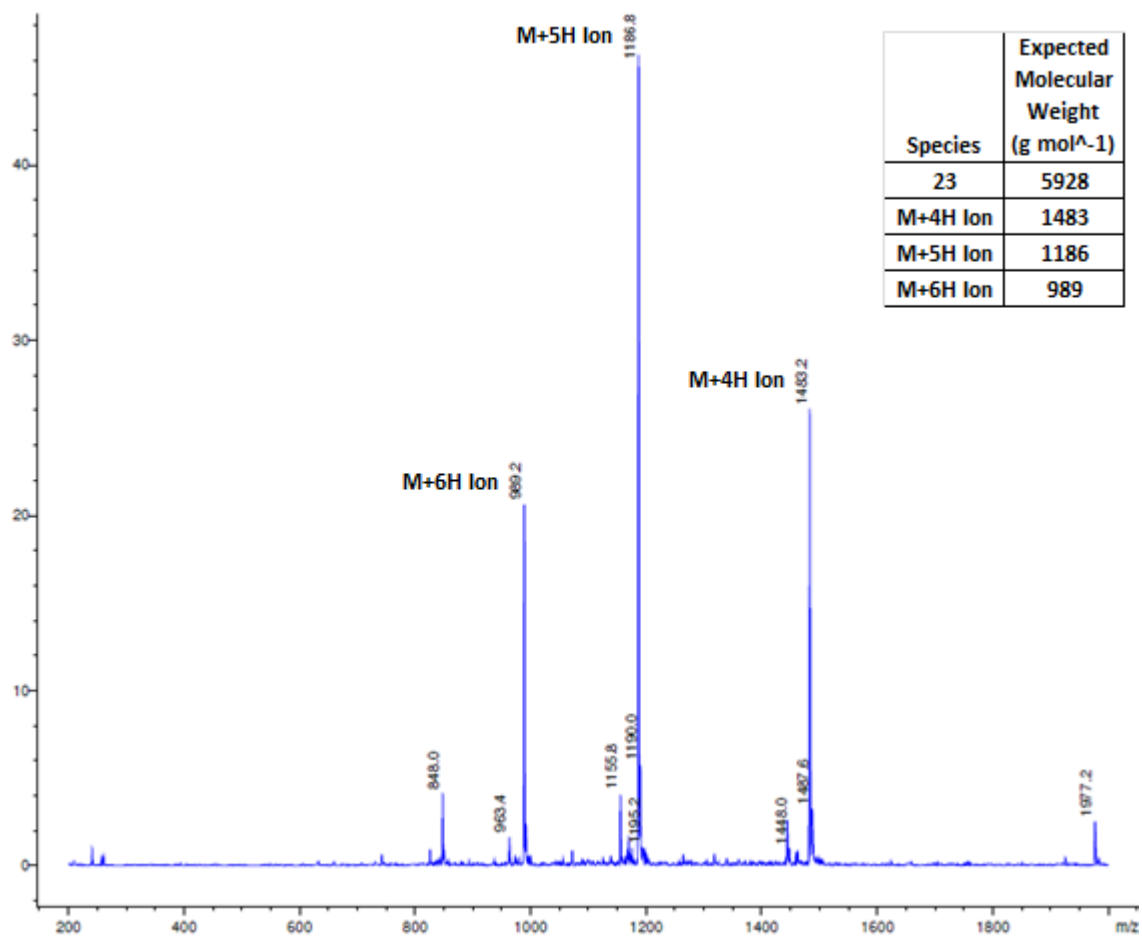
Peak Number	Retention Time (mins)	Area (%)
1	5.36	3.89
2	5.45	91.11

Table A.3. Purity calculation from peak area of LC trace for purified compound 23.

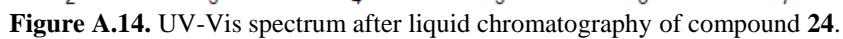
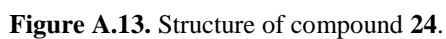
## Appendix



**Figure A.11.** Mass spectrometry detector total ion count recorded during LCMS of purified **23**.

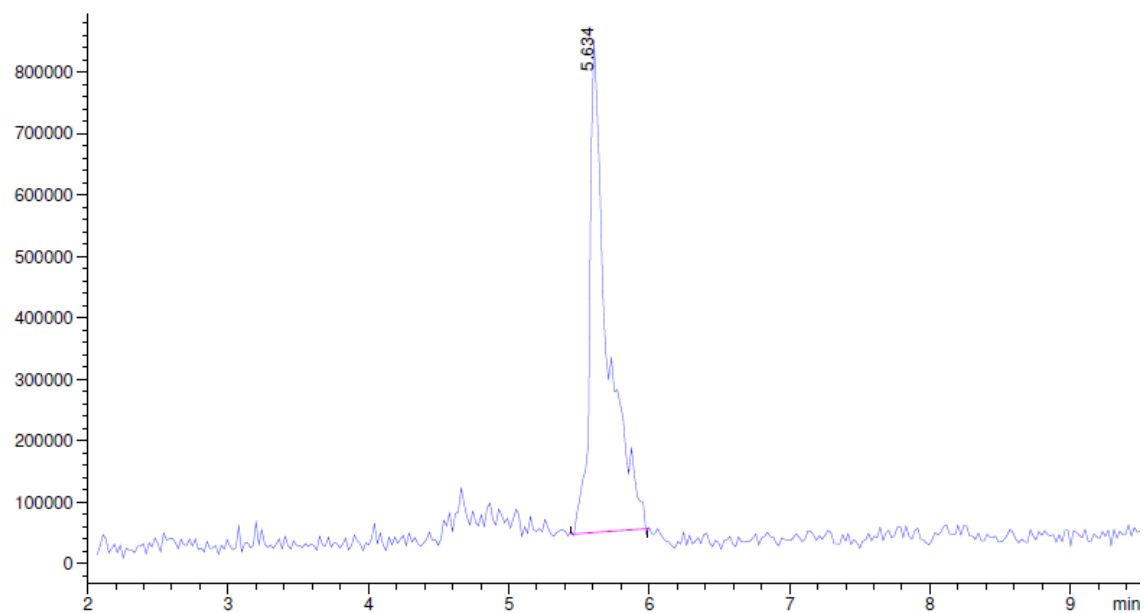


**Figure A.12.** Mass spectrum of peak detected at 5.62 mins.

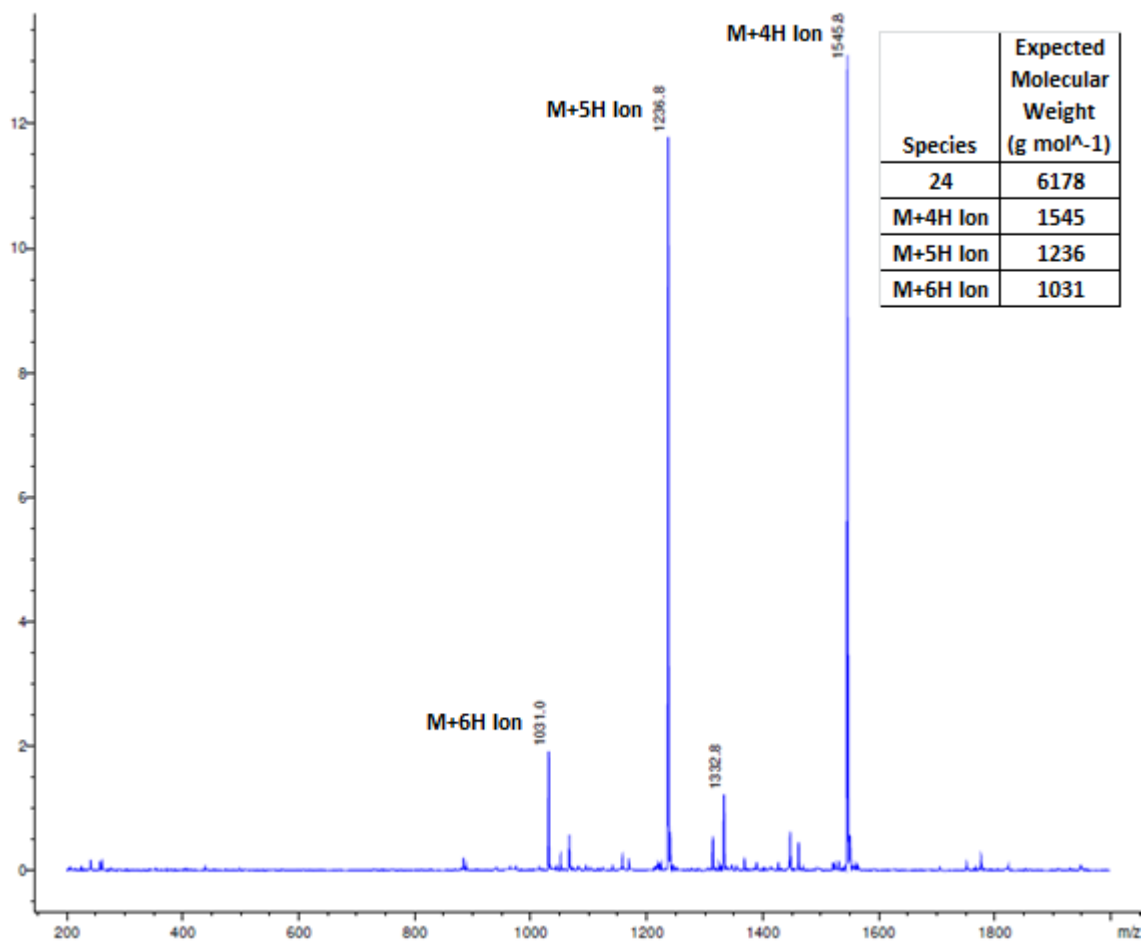


**Table A.4.** Purity calculation from peak area of LC trace for purified compound **24**.

## Appendix



**Figure A.15.** Mass spectrometry detector total ion count recorded during LCMS of purified **24**.



**Figure A.16.** Mass spectrum of peak detected at 5.34 mins.

## A.1.5. Compound 25

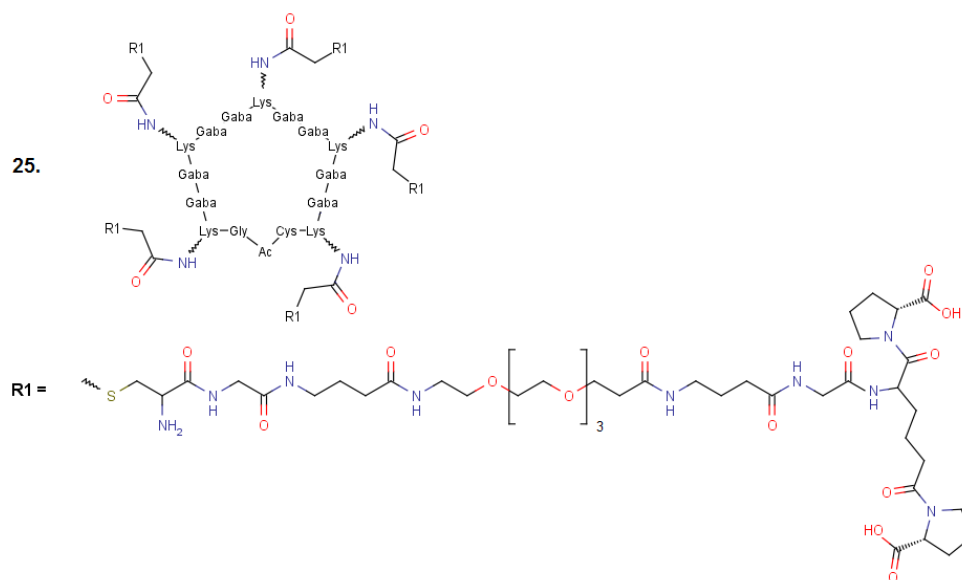


Figure A.17. Structure of compound 25.

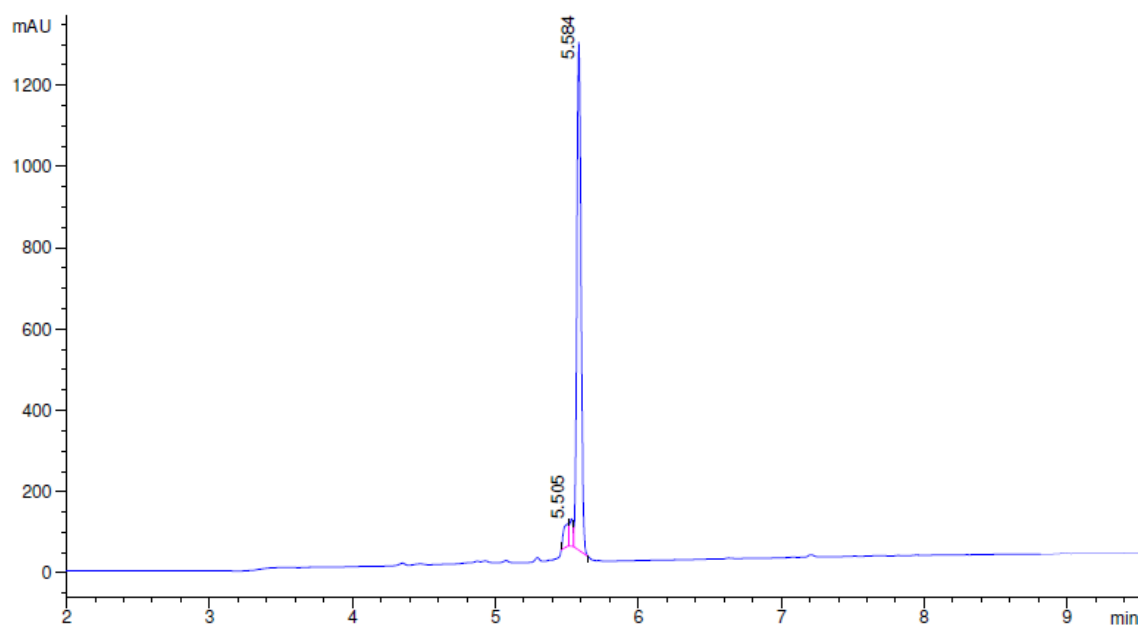
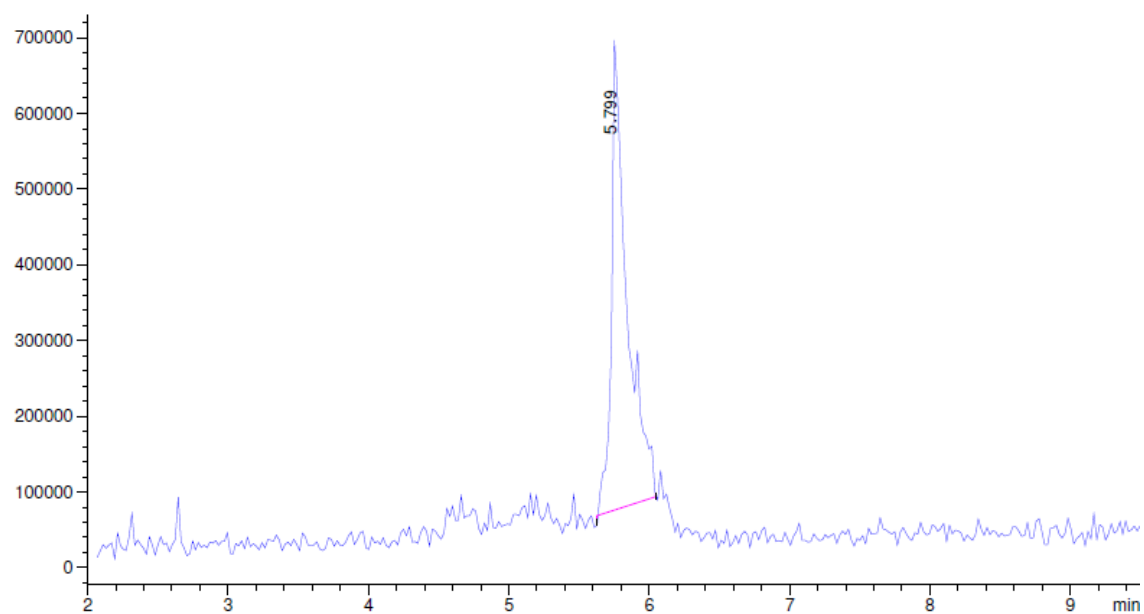


Figure A.18. UV-Vis spectrum after liquid chromatography of compound 25.

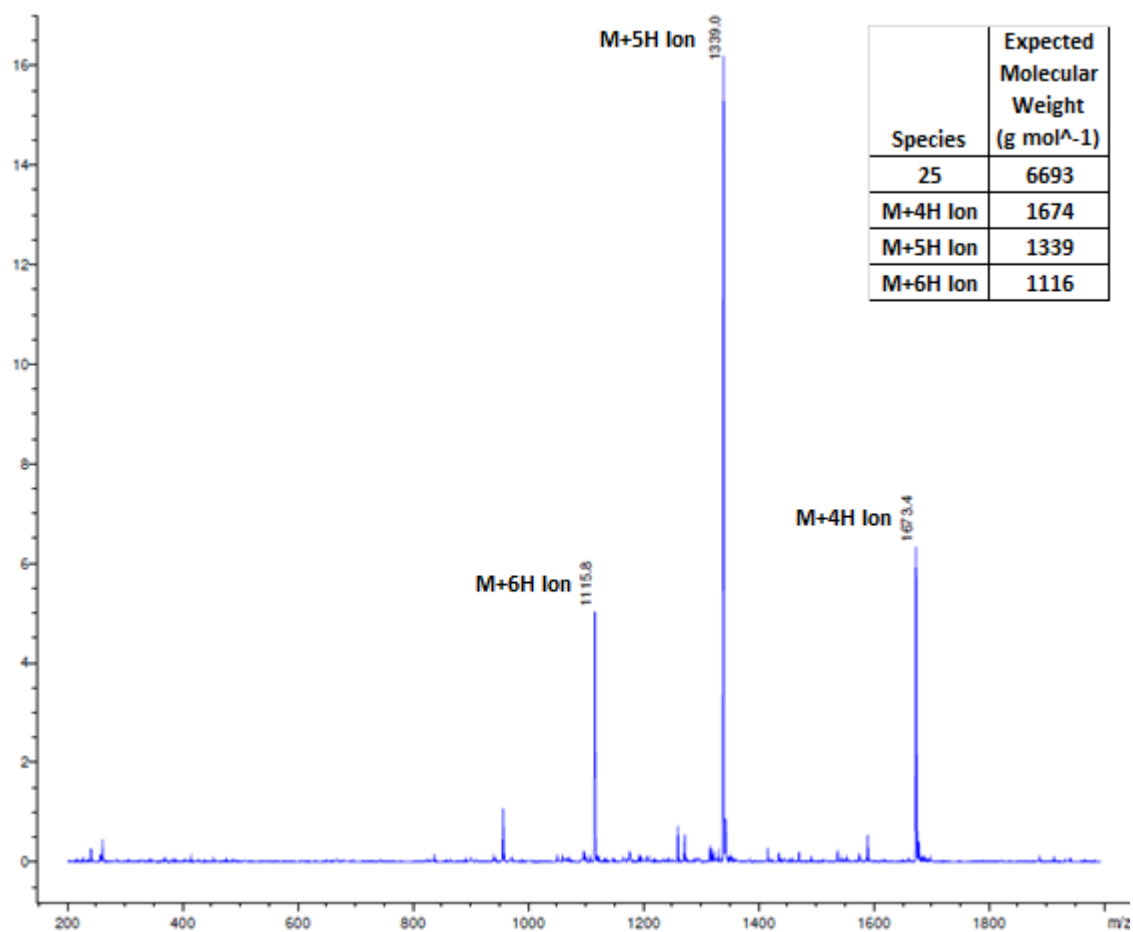
Peak Number	Retention Time (mins)	Area (%)
1	5.51	4.85
2	5.53	4.32
3	5.58	90.83

Table A.5. Purity calculation from peak area of LC trace for purified compound 25

## Appendix



**Figure A.19.** Mass spectrometry detector total ion count recorded during LCMS of purified **25**.



**Figure A.20.** Mass spectrum of peak detected at 5.80 mins.

## A.1.6. Compound 26

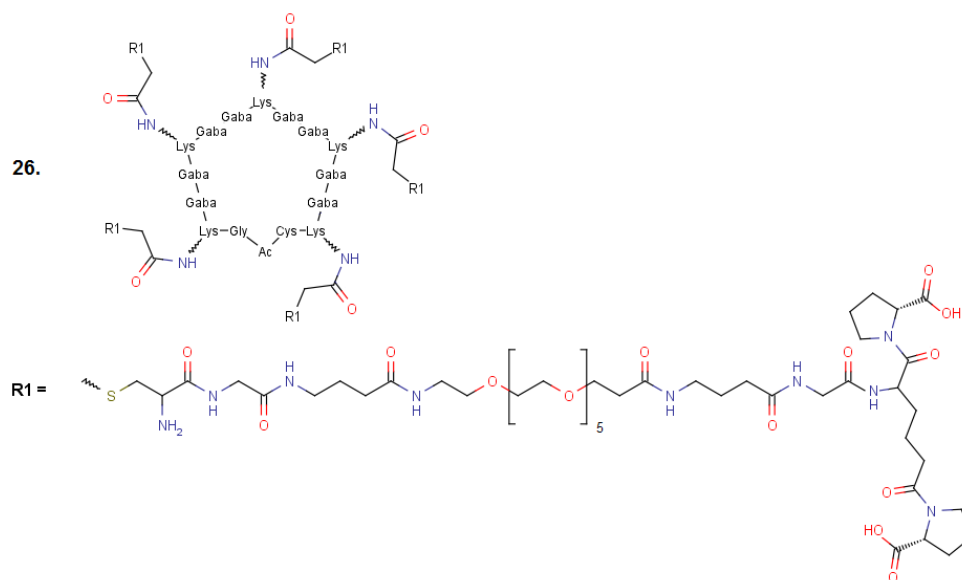


Figure A.21. Structure of compound 26.

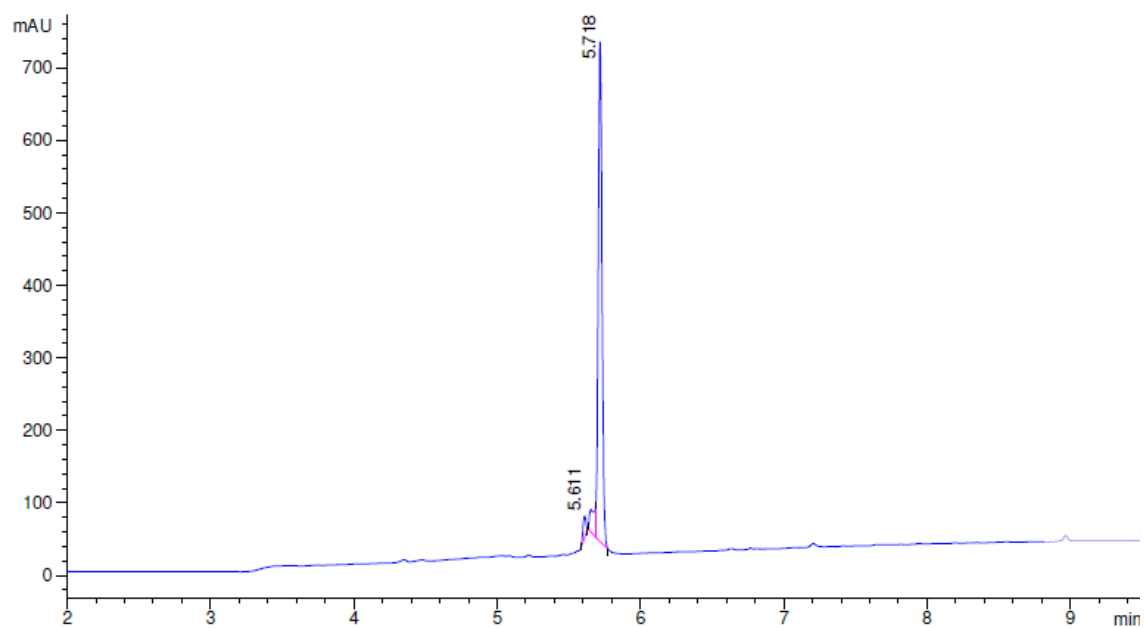
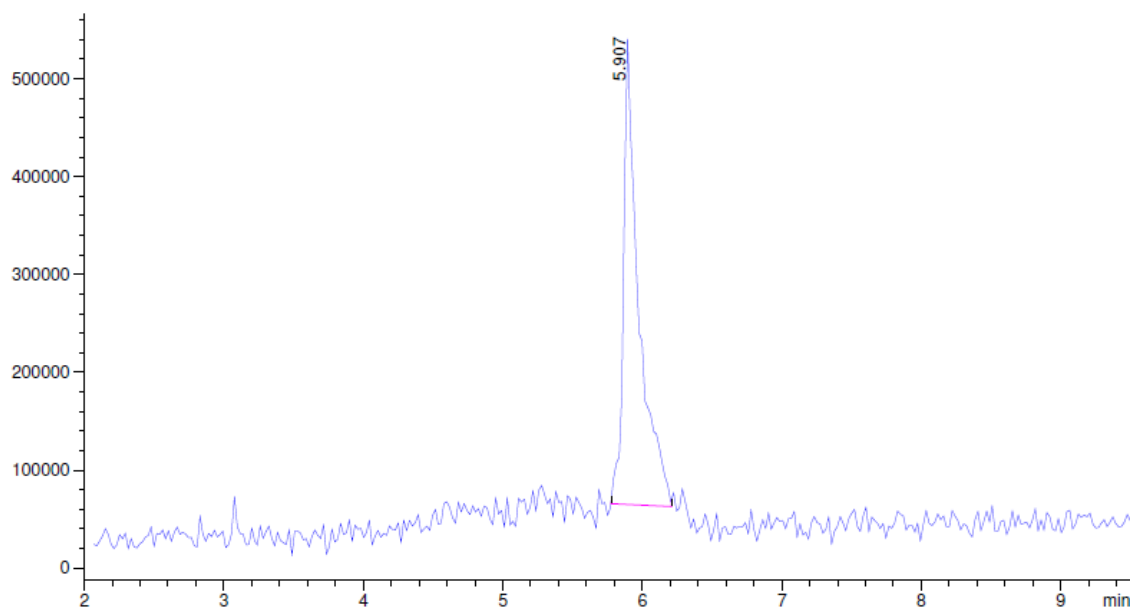


Figure A.22. UV-Vis spectrum after liquid chromatography of compound 26.

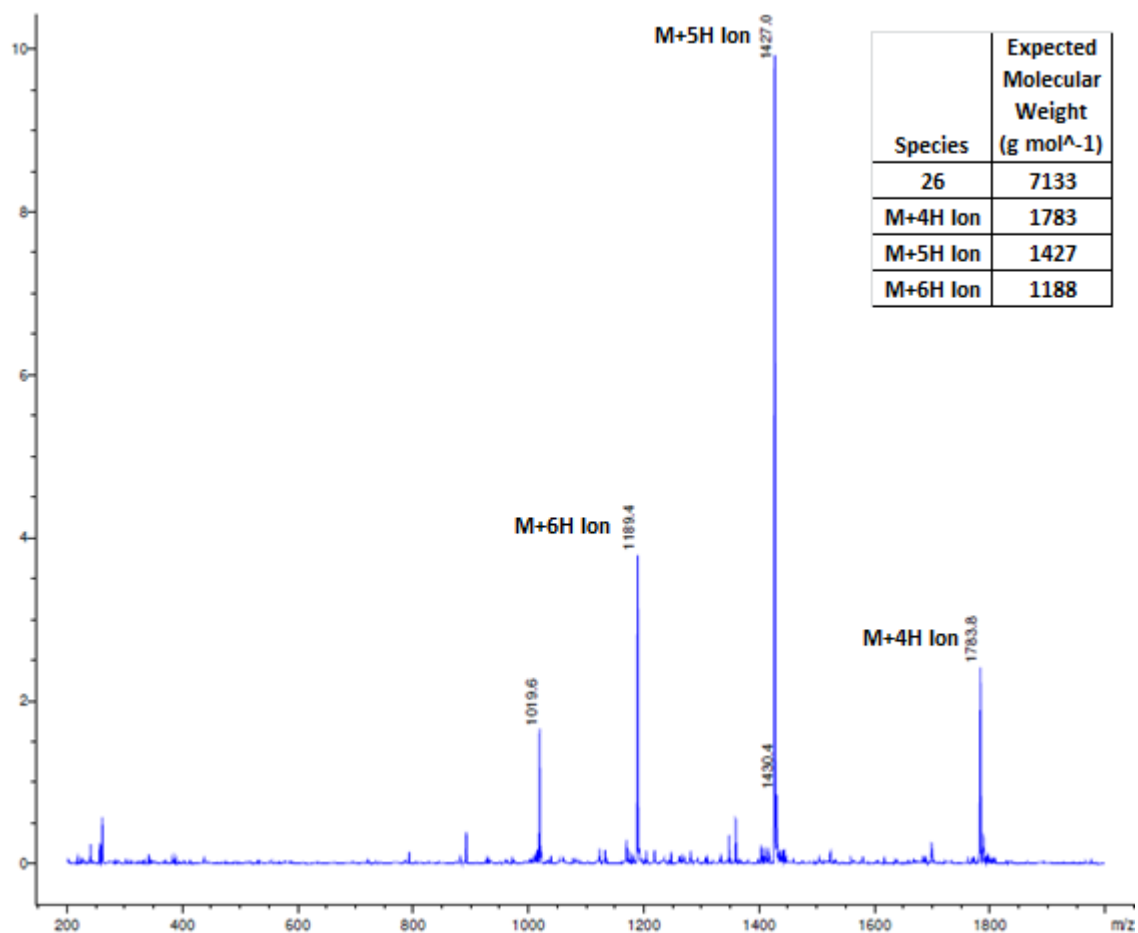
Peak Number	Retention Time (mins)	Area (%)
1	5.61	3.07
2	5.68	6.38
3	5.72	90.55

Table A.6. Purity calculation from peak area of LC trace for purified compound 26.

## Appendix



**Figure A.23.** Mass spectrometry detector total ion count recorded during LCMS of purified **26**.



**Figure A.24.** Mass spectrum of peak detected at 5.91 mins.



## A.2. Mass spectrometry of purified multivalent ligands

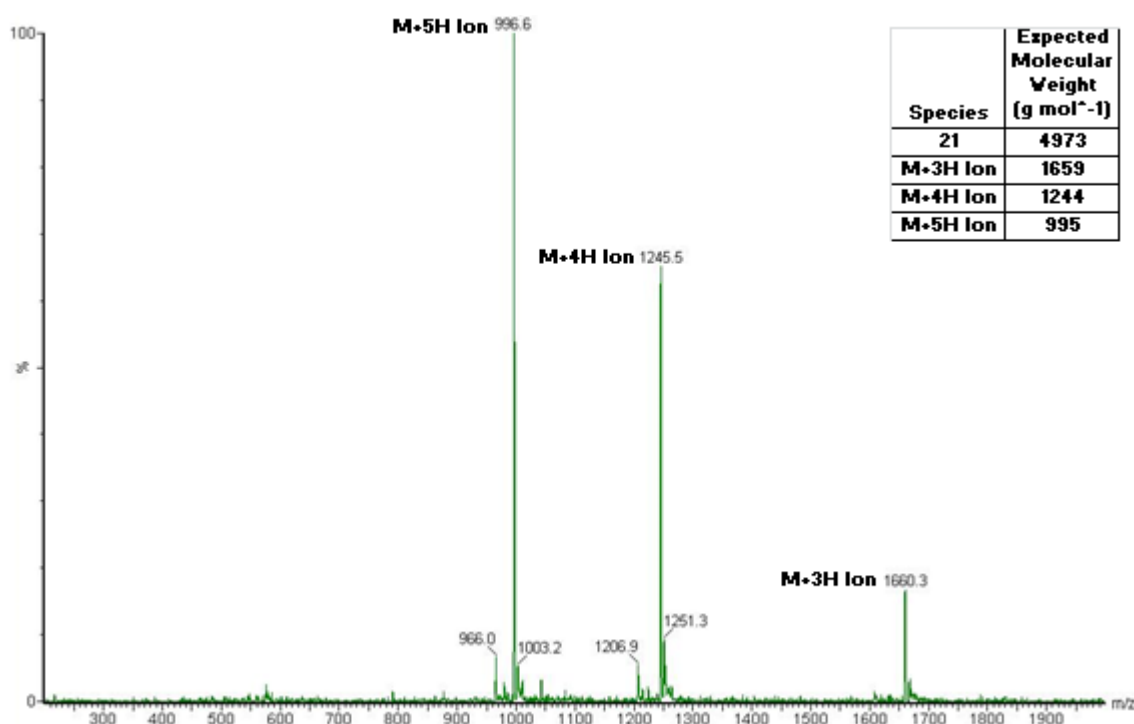


Figure A.25. Mass spectrum of compound 21.

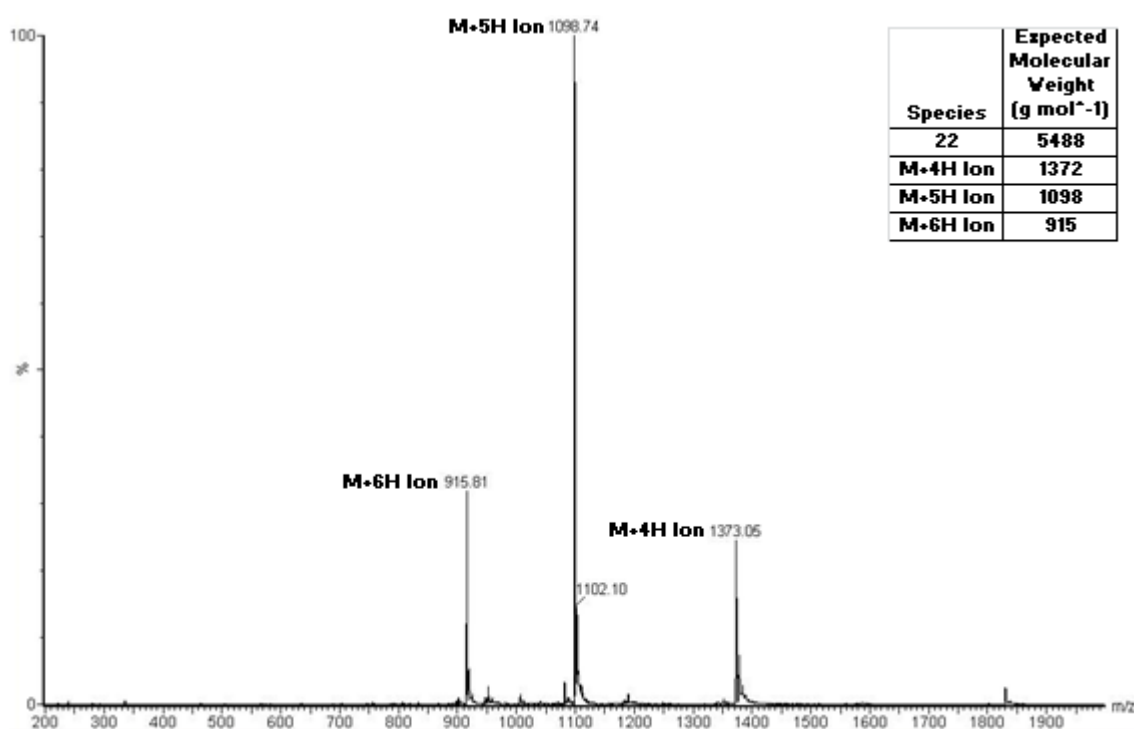


Figure A.26. Mass spectrum of compound 22.

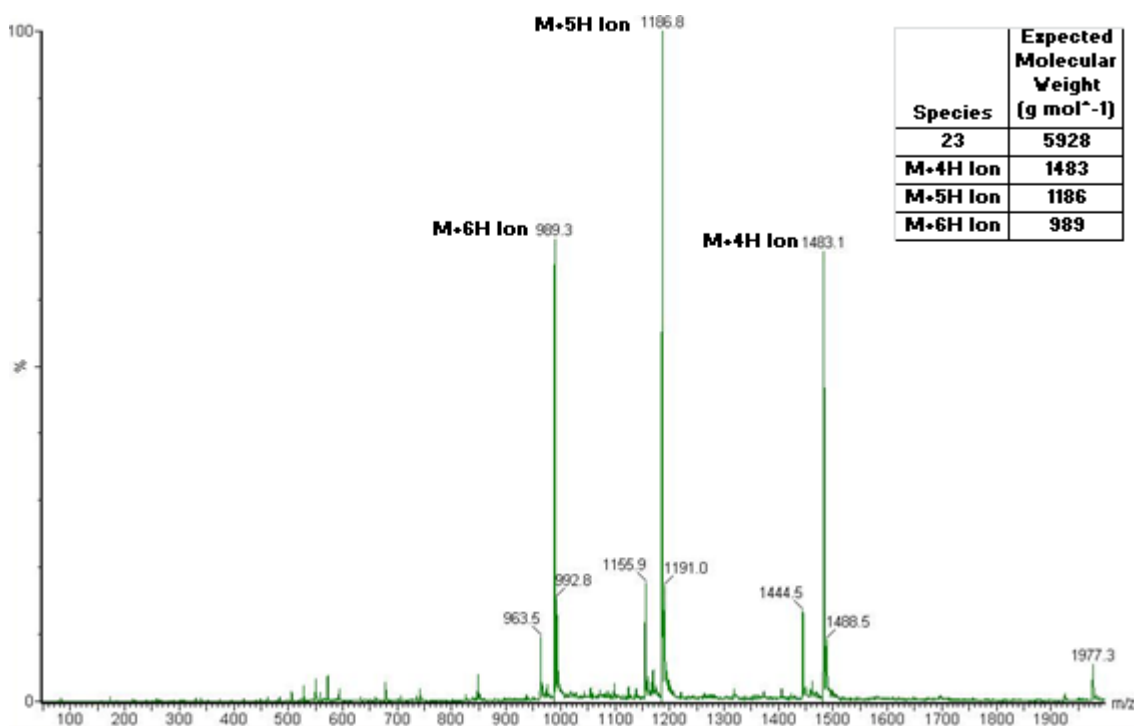


Figure A.27. Mass spectrum of compound 23.

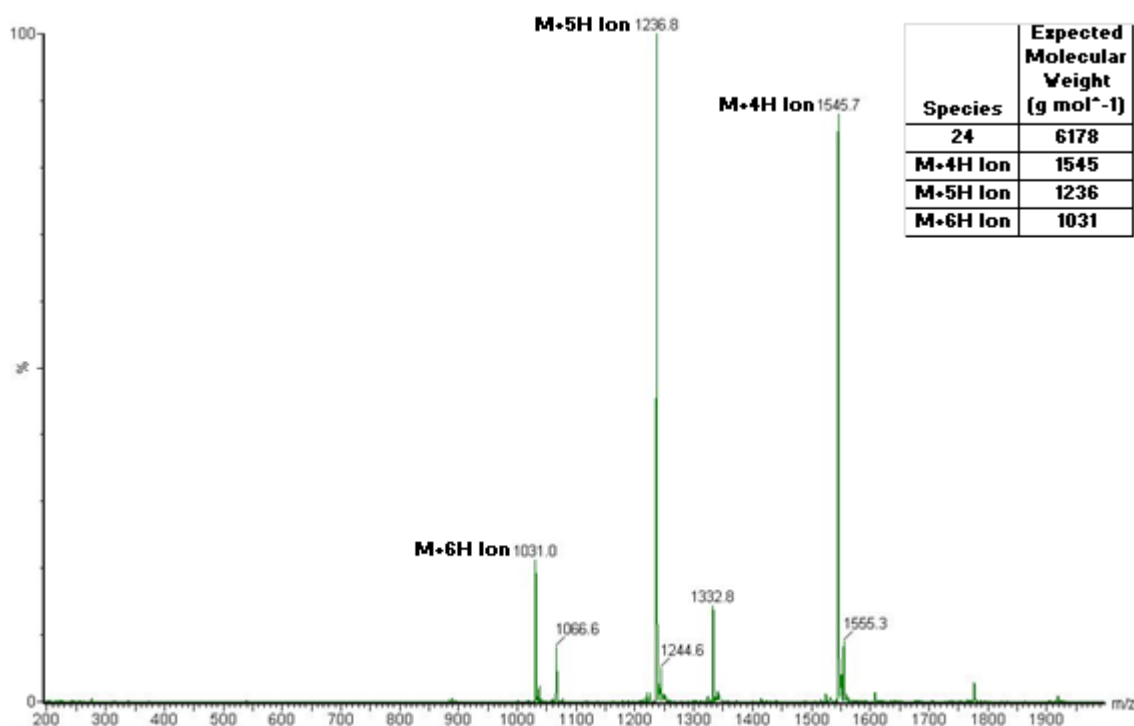


Figure A.28. Mass spectrum of compound 24.

# Appendix

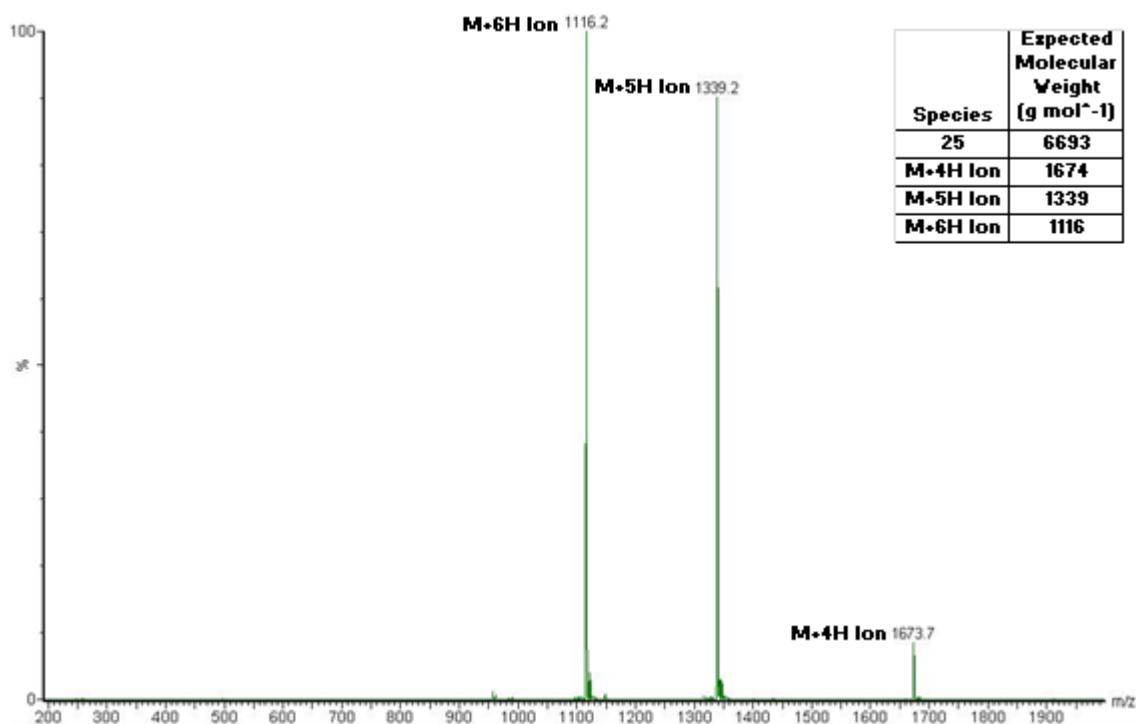


Figure A.29. Mass spectrum of compound 25.

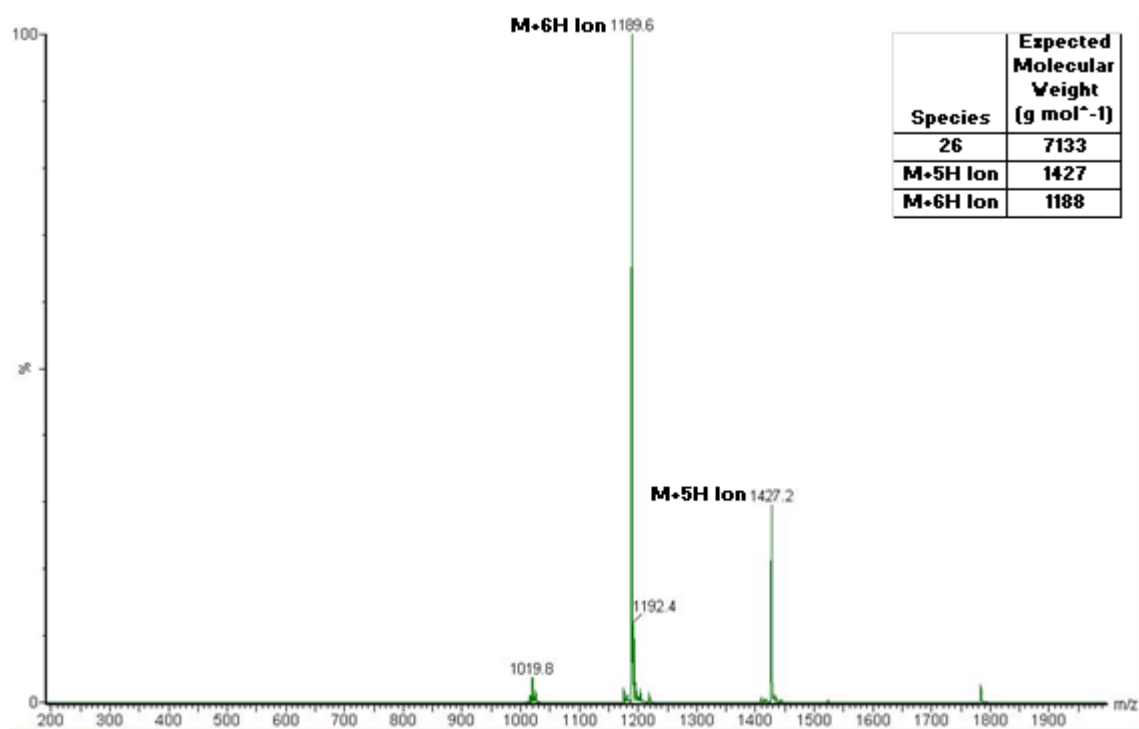


Figure A.30. Mass spectrum of compound 26.

### A.3. Mass spectrometry of purified central scaffold prior to cyclisation (compound 3) and after cyclisation by intramolecular thioether substitution (compound 4)

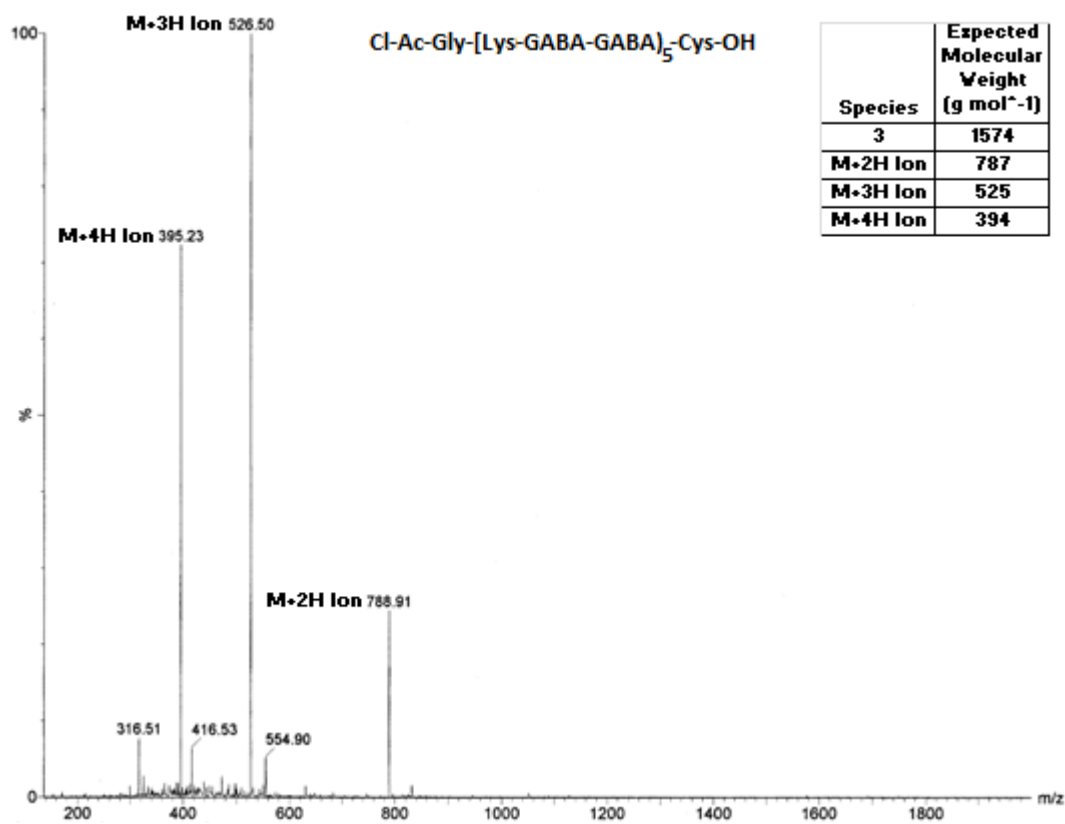


Figure A.31. Uncyclised compound 3 prior to intramolecular thioether substitution reaction.

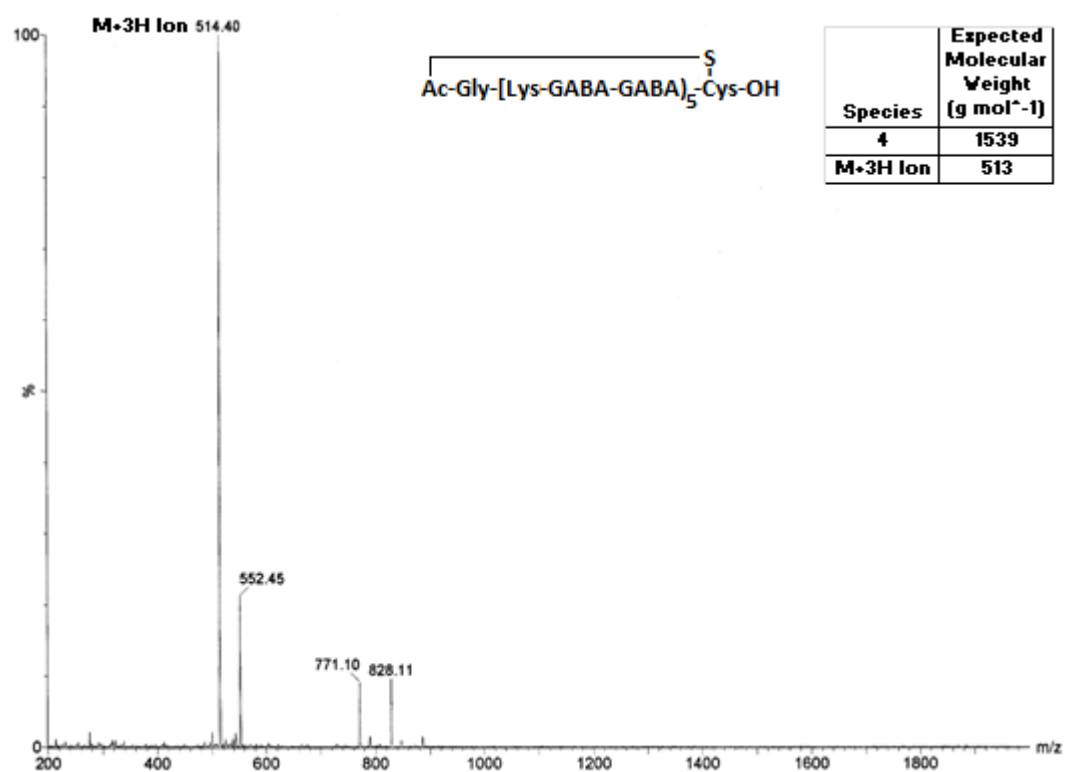


Figure A.32. Cyclic compound 4 after intramolecular thioether substitution reaction and purification.

#### A.4. Derivation of the dissociation constant, $K_d$ , when half macromolecule, $M$ , binding sites are occupied by ligand, $L$ .

The dissociation of constant of the binding event  $[L] + [M] = [ML]$  is equal to:

$$K_d = \frac{[M][L]}{[ML]}$$

$$K_d = \frac{1}{K_a}$$

Where  $K_a$  is the association constant. If  $F_b$  is the fraction of bound macromolecule, then:

$$F_b = \frac{[ML]}{[M] + [L]} = \frac{K[M][L]}{[M] + K[M][L]} = \frac{K[L]}{1 + K[L]}$$

Therefore, the fraction of unbound macromolecule is:

$$1 - F_b = 1 - \frac{K_a[L]}{1 + K_a[L]} = \frac{1}{1 + K_a[L]}$$

Combining the two previous equations gives:

$$\frac{F_b}{1 - F_b} = K_a[L]$$

And when  $F_b = 0.5$ :

$$K_a = \frac{1}{[L]}$$

As  $K_d$  is the reciprocal of  $K_a$ , then:

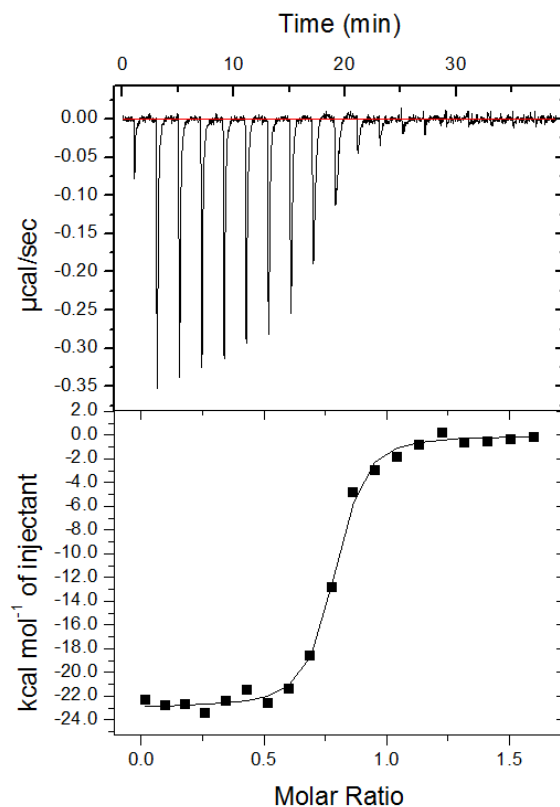
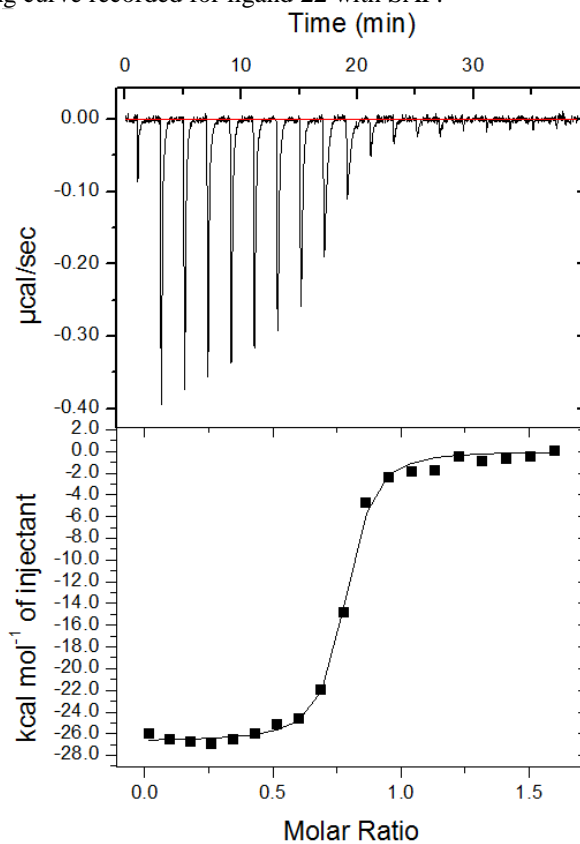
$$\frac{1}{K_a} = K_d = [L]$$

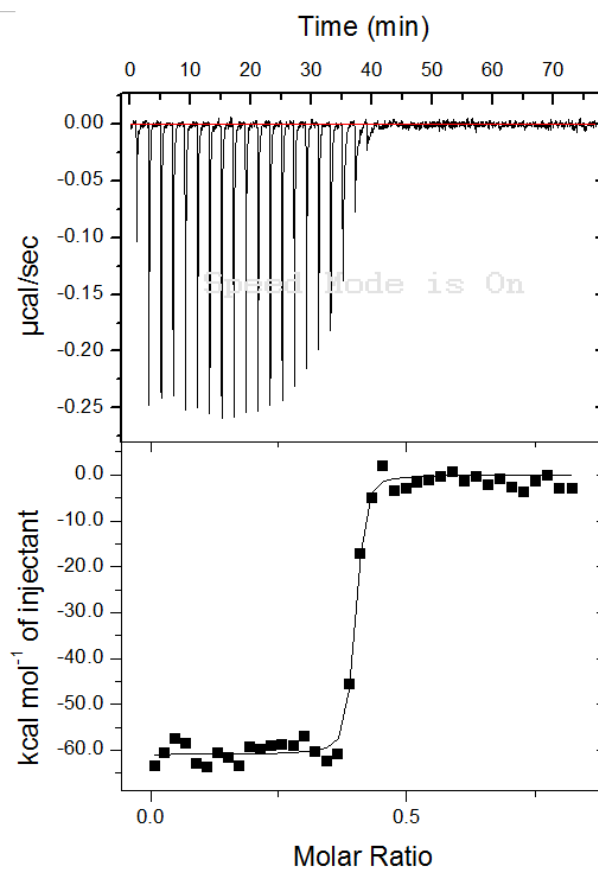
Derivation adapted from Privalov, 2012.

### A.5. Isothermal titration calorimetry results for all runs of multivalent ligands binding with SAP

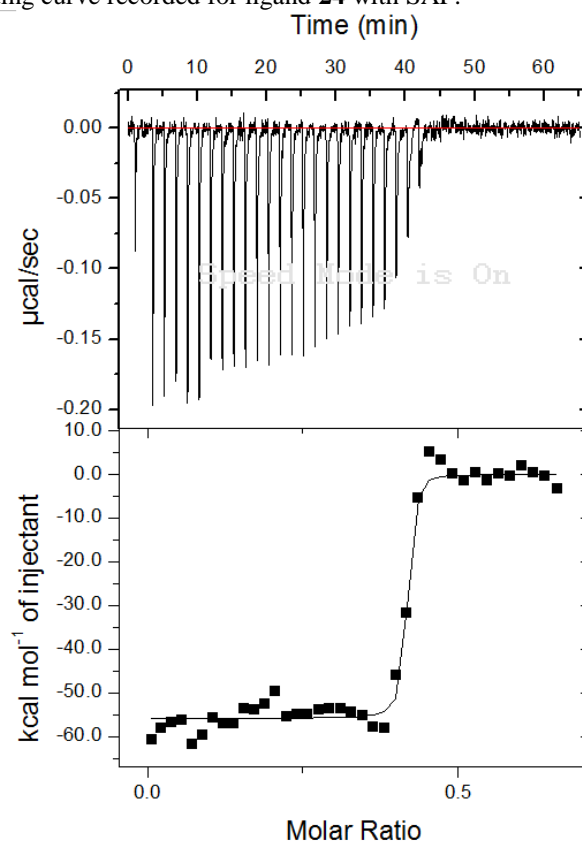
Ligand	Run	K <sub>a</sub> (M <sup>-1</sup> )	K <sub>d</sub> (M)	n	ΔH (cal mol <sup>-1</sup> )	ΔS (cal mol <sup>-1</sup> )
22	1	2.37E7 (±3.54E6)	4.22E-8 (±6.30E-9)	0.751	-2.29E4 (±216)	-43.3
	2	1.87E7 (±2.24E6)	5.35E-8 (±6.41E-9)	0.746	-2.33E4 (±207)	-44.9
	3	6.15E7 (±1.80E7)	1.40E-8 (±4.10E-9)	0.727	-2.35E4 (±301)	-43.0
23	1	1.35E7 (±2.37E6)	7.40E-8 (±1.30E-8)	0.597	-2.54E4 (±436)	-52.6
	2	2.96E7 (±4.36E6)	3.38E-8 (±4.98E-9)	0.745	-2.70E4 (±231)	-55.3
	3	3.79E7 (±1.42E7)	2.64E-8 (±9.89E-9)	0.695	-2.58E4 (±542)	-51.9
24	1	6.28E8 (±1.81E8)	1.59E-9 (±4.58E-10)	0.389	-6.10E4 (±535)	-164
	2	1.00E9 (±3.44E8)	1.00E-10 (±3.44E-11)	0.402	-5.91E4 (±459)	-157
	3	4.59E8 (±1.63E8)	2.18E-9 (±7.74E-10)	0.360	-6.03E4 (±721)	-163
25	1	5.31E9 (±4.28E9)	1.88E-10 (±1.51E-10)	0.414	-5.57E4 (±539)	-142
	2	1.52E9 (±8.29E8)	6.58E-10 (±3.59E-10)	0.410	-6.08E4 (±701)	-163
	3	4.93E9 (±5.08E9)	2.03E-10 (±2.09E-10)	0.413	-5.59E4 (±597)	-146
26	1	1.76E9 (±8.61E8)	5.68E-10 (±2.78E-10)	0.264	-4.33E4 (±362)	-103
	2	1.82E9 (±8.97E8)	5.49E-10 (±2.71E-10)	0.346	-4.49E4 (±562)	-108
	3	7.94E8 (±4.50E8)	1.26E-9 (±7.14E-10)	0.325	-4.68E4 (±539)	-116

**Table A.7.** Isothermal titration calorimetry results for all runs of multivalent ligands **22-26** binding with SAP in the presence of calcium ions.

**A.6. Isothermal titration calorimetry binding curves for multivalent ligands 22-26 with SAP****Figure A.33.** ITC binding curve recorded for ligand **22** with SAP.**Figure A.34.** ITC binding curve recorded for ligand **23** with SAP.

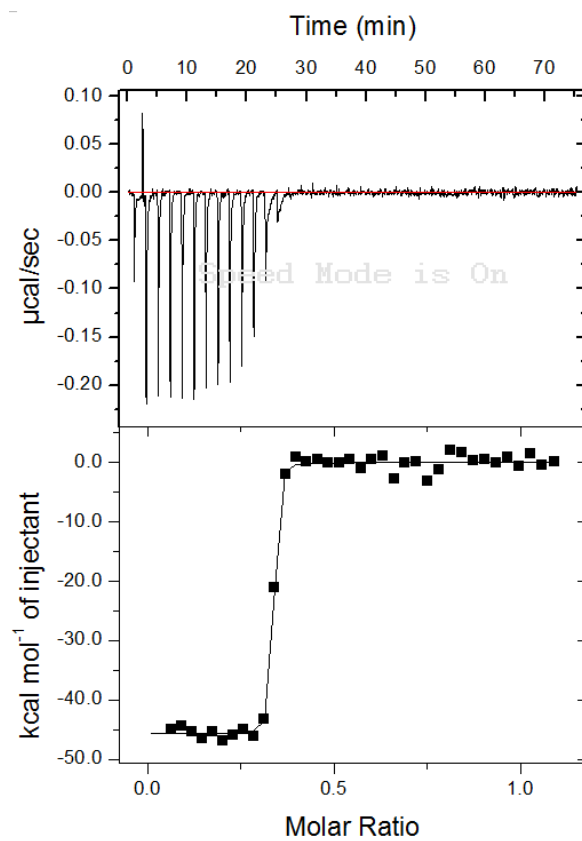


**Figure A.35.** ITC binding curve recorded for ligand **24** with SAP.



**Figure A.36.** ITC binding curve recorded for ligand **25** with SAP.





**Figure A.37.** ITC binding curve recorded for ligand **26** with SAP.